RELATIVE DISTANCE PERCEPTION OF SOUND SOURCES IN CRITICAL LISTENING ENVIRONMENT VIA BINAURAL REPRODUCTION

FOTIS GEORGIOU

School of Computing, Science and Engineering University of Salford, Salford, UK

M.Phil. Thesis 2012

Contents

List of Figures

List of Tables

List of Equations

Abbreviations

BRIR	Binaural Room Impulse Response
D/R	Direct to Reverberant Ratio
HATS	Head and Torso Simulator
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function
IACC	Interaural Cross Correlation
ILD	Interaural Level Difference
ITD	Interaural Time Difference
RMS	Root Mean Square
TOA	Time of Arrival

Dedication

This dissertation is dedicated to my parents, Kaiti Germanou and Apostolos Georgiou. Their support, encouragement, and constant love have sustained me throughout my life.

Acknowledgments

I am most grateful to the project supervisor, Dr. Bruno Fazenda for his time, guidance, and encouragement.

A special thanks is owned to Artemis who was close to me and her support has been precious.

I would like to thank the staff members from the Acoustics Department at the University of Salford for providing me technical support and all the required equipment for the implementation of the project, and all the twenty-three subjects who participated in the listening experiment.

I would also like to thank my aunt Maro who helped me during the most difficult period of the project, my friends Andreas, Trifon, Dimitris, and Manos who encouraged me all these years, and all my friends from Skopelos who always believed in me and helped me to succeed in my goals.

Abstract

Accurate distance cues are important in the degree of realism provided by virtual audio systems. In the last decade there has been an increased interest in this research area. The main focus of this research project is to investigate the effect of different acoustic cues related to distance perception, such as Direct to Reverberant ratio (D/R), in the perception of the relative distance between sound sources in a virtual medium sized critical listening room. The virtual sources were generated by convolving a dry speech signal with modelled and measured BRIRs. The BRIRs were modelled using a direction related image source model for the early reflections and exponentially decaying noise for the reverb tail. In order to investigate relative distance perception and the factors that affect it, a pairwise comparison was conducted involving twenty-three subjects. Three different distances ranging between 1.0m and 3.0m were used in the comparison pairs. The main outcomes from the tests are: 1) Modelled and measured BRIRs provide relative distance cues equally well; 2) Direct-to-reverberant ratio is a significant relative distance cue, even when level between virtual sources is normalized; 3) Adding level differences between the sources does not have a significant effect on the perception of relative distance. However, it reduced the precedence of wrong relative distance judgments by 5%-15%; 4) Manipulation of early reflection time of arrival (TOA) does not appear to be a significant cue in distance perception. These findings are important in the field of virtual reality and computer gaming because they show that the relative distance of a virtual source can be manipulated simply by adjusting the direct-to-reverberant ratio of the BRIRs. It can thus be concluded that large BRIR databases and interpolation between BRIRs at different distances are not required for appropriate distance cues.

1. Introduction

1.1 Research motivation and project objectives

Humans localise sound sources in three-dimensions. The egocentric location of a source is specified in terms of the two parameters: its direction, dependent on azimuth (lateral direction with respect to the facing direction of the head) and elevation (direction with respect to the ear-level plane); and its distance (how far or close is the source from the listener's head).

Distance perception of sound sources has received little attention in comparison to directional localisation and other spatial hearing mechanisms. Distance perception is a more complicated task than directional localisation. The mechanisms and the way people perceive distances are not yet fully understood. However, over the past fifteen years there has been an increased interest in this research area and researchers have managed to identify many acoustic and non-acoustic cues that are related to distance perception.

There are two factors in distance perception, the egocentric and the exocentric. Exocentric or relative factors influence the differences in distance perception of the relative positions between sound sources and do not themselves provide absolute distance information unless the listeners have previous information about the characteristics of the source. Egocentric or absolute cues, on the other hand, are the variables that provide information about the absolute location of the sound source. Also, egocentric distance is the apparent distance of a sound source and exocentric or relative distance is referred to as the relative distance between sound sources. Basically, the "relative" part means that there is some external reference to the listener (Mershon 1979, Zahorik 2002a, Shinn-Cunningham 2000b, Nielsen 1993).

A very obvious acoustic cue for source distance is a change in the overall level with distance (greater level when source is near and smaller when source is far). Another strong distance cue is the ratio between direct-and-reverberant energy (D/R). D/R is the ratio of the energy reaching the human ears directly from the source over the energy arriving at the ears via one or more reflections. The source level is a relative distance cue and cannot, by itself, provide information about the absolute source distance unless the listener has previous information about the characteristics of the sound source (Mershon 1975). Conversely, D/R is a strong absolute distance cue (Zahorik 2002a, 2000c; Larsen 2008). Interaural level differences (level difference between signals arriving at the left and right ear due to the shadowing effect of the head and the extra path that the sound has to travel to reach the farther ear relative to the sound source) can provide also distance information for sources outside the median plane. Interaural level difference can serve both as relative and absolute distance cues for near distances (Fukuda 2003; Brungart 1999). Other acoustic distance cues are the change in spectrum of the source due to air absorption; and low frequency absorption of the materials of a room (Butler 1980; Blauert 1976). Two non-acoustics cues that strongly affect distance perception are the familiarity (how familiar the listener is with the listening environment and the source signal) and vision (Devalez 2008; Zahorik 2001; Mershon 1980). Vision plays a primary role in distance perception, but visual information cannot always be provided to the listener. For example, in the work presented here listeners were listening to binaural reproduced sound sources but they were not given any visual information either for the source or the modelled environment.

When the source is synthesized binaurally, the information about the source location within the room is provided from the binaural room impulse response (BRIR). The BRIR describes the sound transfer from the source to the receiver ears. The BRIR includes the effects of the sound reflections from boundaries and the effects of the diffractions on the listener's head and body including the microreflections of the pinna. Therefore, BRIR's are very important on the auralisation techniques such as binaural reproduction through headphones.

Several studies have looked into distance perception via binaural reproduction, which shows that this technique reproduces distance well. Some of these studies are briefly listed in this paragraph. Zahorik (2002a, 1997) focused on the absolute distance perception in reverberant environments and examined how people weight the two main distance cues (source level and D/R) under different angular locations, distances, and source signals. Zahorik (2002c) and Larsen (2008) investigated the D/R sensitivity. Kim (2001) and Shinn-Cunningham (2000b) looked into the effect of binaural cues in distance perception for nearby sources. Shinn-Cunningham (2000b) examined distance perception under reverberant and anechoic conditions for nearby sources. Fukuda (2003) researched the effect of ILD in relative distance perception. Other researchers investigated the effect of non-acoustic cues, such as vision and learning, in distance perception (Mershon 1980; Zahorik 2001; Schoolmaster 2003a).

In this project BRIRs were modelled and measured in a critical listening room with dimensions 6.6m×5.8m×2.8m and reverberation time of 0.27 seconds. These BRIRs were used to investigate the project's main objectives which are listed below:

1. Investigate the effect of D/R on relative distance perception.

2. Evaluate differences between modelled and measured BRIRs in providing relative distance cues.

3. Examine the effect of early reflections TOA on relative distance perception.

1.2 Methodology

One pairwise listening test was performed which was separated in three parts. Subjects were asked to indicate in a continuous dimensionless scale how far or close they perceive the second sound of the pair in comparison to the first sound. The source signal was a 4 seconds long Italian speech sample.

In the first part of the listening test, relative distance perception was evaluated both for measured and modelled BRIRs and overall level normalized. In the second part the effect of D/R and reflections TOA in relative distance perception was investigated. The third part was exactly the same as the first but the overall level between virtual sources was not normalized. The results were analysed by a non-parametric ANOVA called Kruskal–Wallis.

2. Basic Theory

In this section basic theory that is directly related to the project is briefly demonstrated. This section includes aspects on binaural cues, including the concept of binaural synthesis and room acoustics.

2.1 Head and Torso related cues

The three main head related cues are the interaural time difference (ITD), the interaural level difference (ILD), and the head and torso spectral cues. Interaural time difference is a consequence of the physical separation of the two ears (figure 2.1). ITD occurs due to the extra path that the sound wave has to travel to reach the ear that is farther from the sound source. Interaural level difference (ILD) refers to the amplitude differences between the two ears, which occurs due to the shadowing effect of the head when the source moves away from the median plane (Blauert 1976; Howard and Angus 2009).

ITD is the dominant localisation cue for low frequencies where the sound wavelengths are much larger than the diameter of head. ILD is the dominant localisation cue for high frequencies where the sound wavelength becomes smaller than the diameter of the head. There is a crossover between the two frequency related cues between 700Hz and 2.8kHz, which makes our ability to localise sound at that frequency range not as good in comparison to other frequencies (Howard and Angus 2009).

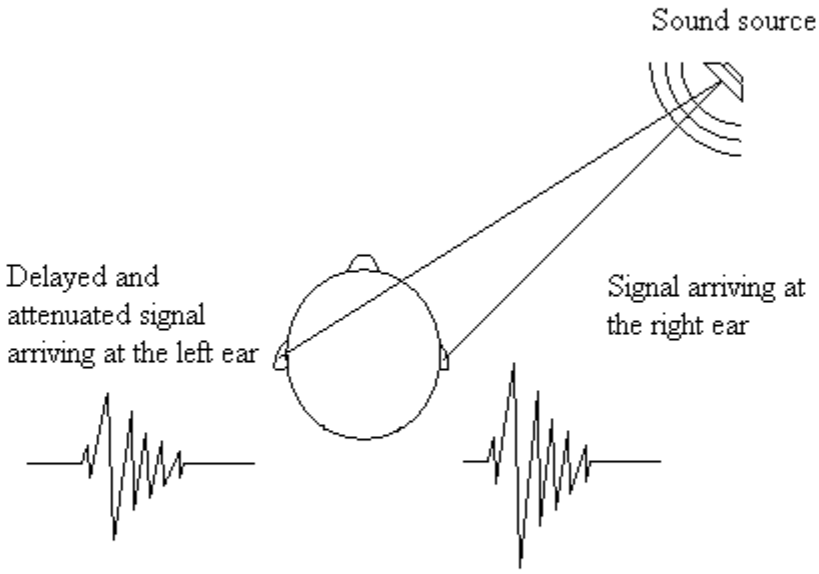


Figure 2.1: Interaural Time Difference (ITD) and Interaural Level Difference (ILD)

Spectral cues occur due to the reflections at the outer ear (pinna), torso and head. The complex construction of the outer ear causes a unique set of micro delays, diffractions, and resonances that form a comb filter effect. These effects occur typically at frequencies above 5kHz. Spectral cues are the main cues for front-back and elevation localisation and are unique for every person (Møller 1995). The combination of these three cues is known as Head-Related Transfer Function (HRTF). HRTFs have a very important role in 3D sound reproduction.

These three head related cues, and especially the spectral cues, have singular characteristics for each source location. Consequently, there is a singular HRTF for each sound source position. In the time domain, the HRTFs are referred to as Head Related Impulse Responses (HRIR). The head related transfer function combined with the reflections from boundaries and other obstacles is referred as Binaural Room Impulse Response (BRIR).

Further to static cues, humans (and other animals) resolve localisation ambiguities by moving their heads slightly from side-to-side. This head movement changes the location of the sound source relative to ear position, providing a real time manipulation of the three cues described above and allowing a disambiguation of the source direction (Wersenyi G. 2008; Angus and Howard 2009).

2.3 Head and Torso Simulator and Binaural Synthesis

2.3.1 Head and Torso Simulator

Measurement of HRTFs can be made near the eardrum, within the ear canal, or at the entrance of the ear canal (blocked). Møller (1995) argues that blocked ear canal measurements vary less across individuals compared to open ear canal measurements.

Nonetheless, measuring individual HRTFs is very time consuming and in most cases, measurements cannot be made at the listener's own ears (e.g. consumer applications for video game audio). For this purpose more generalized HRTFs measured with Head and Torso simulators (HATS) are used. HATS have constructed features of the head, pinna, and torso of an average human and use fixed microphones at the entrance or inside the ear canals to obtain the pressure response at the measuring points.

2.3.2 Binaural synthesis

If a sound is recorded at the ears of a person and then this recording is played back to that listener through headphones, a complete auditory experience will occur (Carline 1996). The same results can be achieved with the use of Binaural Synthesis. This technique is based on the idea that any sound source can be modelled simply by filtering dry signals with the left and right ear HRTFs corresponding to the desirable virtual source location (Figure 2.2). Thus, in theory, two audio channels are enough to provide a 3D sound experience.

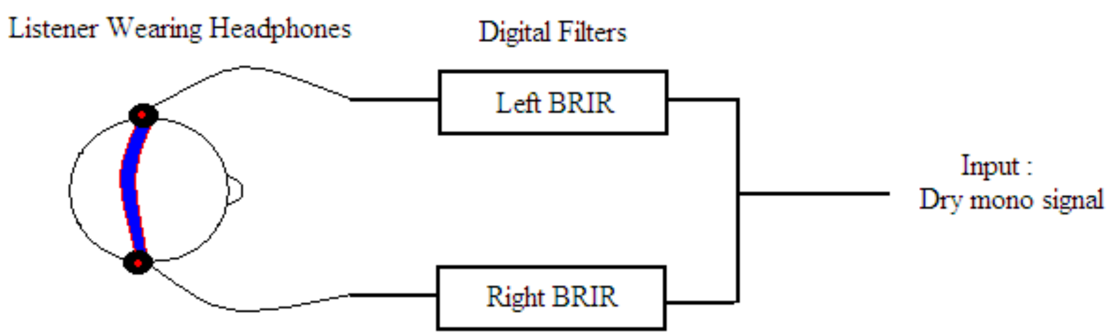


Figure 2.2: Binaural synthesis

Binaural synthesis allows plenty of virtual sound sources to be synthesized simultaneously. Therefore, many virtual tools based on binaural synthesis can be designed to take advantage of this principle. Some examples are virtual monitoring studios, virtual sound systems, and audiovisual interactive software. The quality of such a system is based on its directional and distance localisation performance, on the room model (how realistic it sounds), on the quality of the head tracking system and on how well it can work in real time.

2.4 Room Impulse Response and the importance of its anatomy

Impulse response is the pressure-time response function at the receiver position inside a room as a result of an impulse excitation (Figure 2.3).

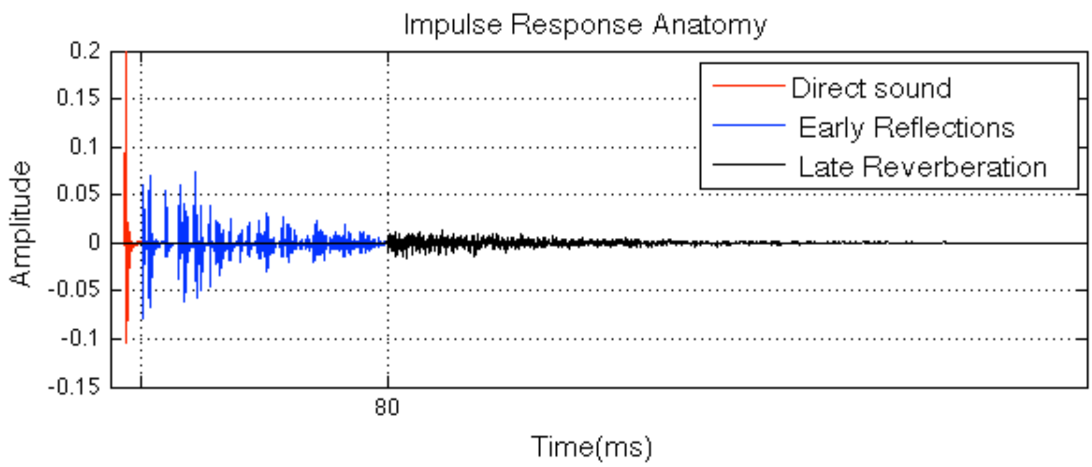


Figure 2.3: Room Impulse Response (Red: direct sound, Blue: early reflections, Black: late reverberation)

The three main parts of the impulse response are:

The Direct sound: The direct sound dominates in the aspect of localisation of the source due to precedence effect.

Early reflections: Strong and low-density reflections within a short TOA (approximately 80ms to 100ms or 15ms to 20ms for small critical listening spaces). The early reflections play an important role in sound perception. Some important aspects of early reflections are listed below. More information can be found in Howard and Angus (1999) and Haas (1951).

- Precedence effect: When the same source arrives within 30ms delayed relative to the direct sound due to a reflection off a room boundary only the direction of the direct sound is perceived (Wilslon K. 2005).
- Summing localisation: If a reflection arrives within 2ms after the direct sound, it is grouped together with the direct sound and causes an image shift (the direction of the sound source is perceived somewhere between the early reflection and the direct sound) (Blauert 1976).
- If the initial time gap between the first reflection and the direct sound is long, the reflection is separated from the direct sound and is perceived as an echo.
- If strong reflections arrive at short delay time (approximately 20ms), comb filtering effect occurs which can cause coloration on the perceived sound
- Reflections arriving within 50ms to 100ms support the direct sound. Reflections up to 50ms are important for speech intelligibly as they increase the amount of speech energy arriving at the listener. Reflections up to 80ms are important for music clarity.
- First order reflections are also important for the apparent source width and listener envelopment.

Late reverberation: After approximately 80ms the reflection density increases rapidly forming the reverberation tail. Early energy decays quickly and late energy decays exponentially (Figure 2.3). Reverberation is frequency dependant due to absorption characteristics of the materials inside the room and air absorption. Therefore, different frequency bands have different decaying times. The late reverberation is spread uniformly around the room and in an ideal field it may be considered diffuse.

For the implementation of this project BRIRs were modelled. The anatomy of the BRIRs is exactly the same as the anatomy of the impulse responses. The only difference between them is that BRIRs are specifically made up of the head related impulse responses (HRIRs) combined with the room response.

2.6 Conclusion

In this section basic theory about head related transfer function, anatomy of room impulse responses, effect of early reflection and reverberation, binaural synthesis, were discussed.

3. Perception of distance

In this section the literature review related to the main subject of this dissertation is addressed. It covers the topics of distance perception and externalisation of sound sources in headphone listening.

3.1 Distance perception cues

3.1.1 Source level

The level of a sound source can provide a strong distance cue. The level of a source decreases with distance and thus sources with lower level tend to be perceived farther away than those with higher level (Mershon 1975; Gardner 1968). In free field environments and for distances above 1.0m, the sound level follows an inverse square law where the level of the sound source is proportional to 1/r (r: distance in metres). This implies a 6dB reduction for every doubling of distance (Equation 3.1). The above relationship is applicable only for outdoors sound propagation. Indoors, the reflected energy from the room boundaries affects the way sound pressure level decreases with distance in such a way that the inverse square relationship is no longer maintained. In this situation the perception of distance is no longer directly correlated with source level but with the D/R, which is discussed in subsection 3.1.2 below.

$$\Delta L = 20 \times log_{10}\left(\frac{r_1}{r_2}\right)$$

Equation 3.1: Inverse square law

Where:
r_1: Distance (in metres) between receiver and first sound source
r_2: Distance (in metres) between receiver and second sound source
ΔL: Level difference between two identical sound sources at r2 and r1

In the listener's near field (below 1.0m), the direction of the source affects distance perception. As has been indicated by Shinn-Cunningham (2000b), at such short distances the level of the source relative to the ears does no longer obey the inverse square law, and distance perception becomes direction dependent (Figure 3.1).

Figure 3.1 shows how the root mean square level (RMS) at the near ear varies with source distance and angle. The level at the near ear decreases faster at closer distances when the source is located at 90° compared to the source at 0°. For example, the level difference at the near ear between source at 0.2m and 1.0m is much larger when the source is located at 90° than at 0°. Since a very strong relative distance cue is different in the two angle orientations, distance perception becomes angle depended. It is also clear that the lines presenting the RMS level at the near ear over distance have the same slope above 1.0m, and level cue is no longer angle dependent.

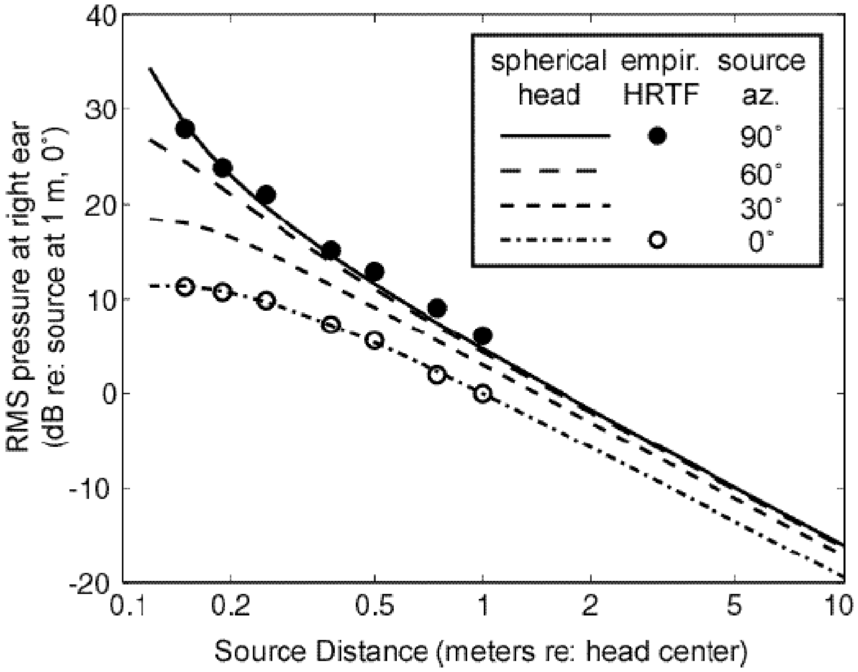


Figure 3.1: RMS pressure at the near ear as a function of source distance (relative to the centre of the head) for sources at various directions in the horizontal plane. Lines show predictions for a perfectly rigid, spherical head. Symbols show measured RMS pressure at the ear canal of a human subject. (Shinn-Cunningham 2000b)

Humans are very sensitive to level differences. It is argued that the smallest perceptible distance change of a median plane source corresponds to the smallest perceptible change in the overall level of a broadband source which is 0.4dB (Strybel 1984).

Despite its relevance for source location perception, source level is an ambiguous cue because it is dependent on intrinsic alterations on the source level itself; these alterations are not dependent on source displacement. As such, level is mainly reliable only as a relative cue for distance estimation unless the listener has a previous experience about the level of the source (Zahorik 2005; Mershon 1975).

In terms of perception, the sensation that humans perceive is loudness and not sound pressure level. Loudness is basically the level of sound as perceived by each individual. The loudness of a sound can depend on a variety of factors in addition to level differences caused by physical distance. Loudness does not vary linearly with level while it also depends on the spectrum and the duration of the sound (Howard and Angus 2009). Loudness of speech in reverberant environments is based both on the D/R (see subsection 3.2.2) as well as on the level of the direct and reverberant components (Warren 1973).

3.1.2 Reverberation and direct to reverberant ratio

In anechoic conditions, absolute distance judgments can be made for sources at distances up to 1.0m (section 3.1.3) and level can only serve as a relative distance cue. In natural (reverberant) conditions other cues are provided which underlie distance perception both for near and far sources.

Mershon (1975) investigated the effect of reverberation in distance perception. He conducted the experiment under anechoic and reverberant conditions. The source was a loudspeaker generating 5 seconds long white noise and the subjects were asked to judge the apparent distance of the different sound sources. The result shows that in reverberant environments judgments were much more accurate. This is also supported (for nearby sources) by Shinn-Cunningham (2005) and Santarelli (2001). Mershon concluded that reverberation may serve both as an absolute and as a relative distance cue.

A dominant cue for distance perception is the Direct-to-reverberant ratio (D/R) (Zahorik 2002, 2005; Larsen 2008; Nielsen 1993). The direct sound always follows an inverse square law regardless of whether the propagation takes place indoors or outdoors (although this assumption is invalid for distances below 1.0m as described in subsection 3.1.1). The level of reverberant energy remains almost independent from source distance for distances above the critical distance (Critical distance is the distance from a source at which the direct and reverberant sound levels are equal). Therefore, in a room, the D/R is inversely proportional to the source distance and may be used as an absolute cue for distance perception. Reverberation is also produced in outdoor environments due to reflection from trees, buildings etc, so D/R can also vary in some outdoor conditions (Zahorik 2005). D/R may be considered an absolute cue, providing distance information with only one presentation of the sound source (Zahorik 2002a).

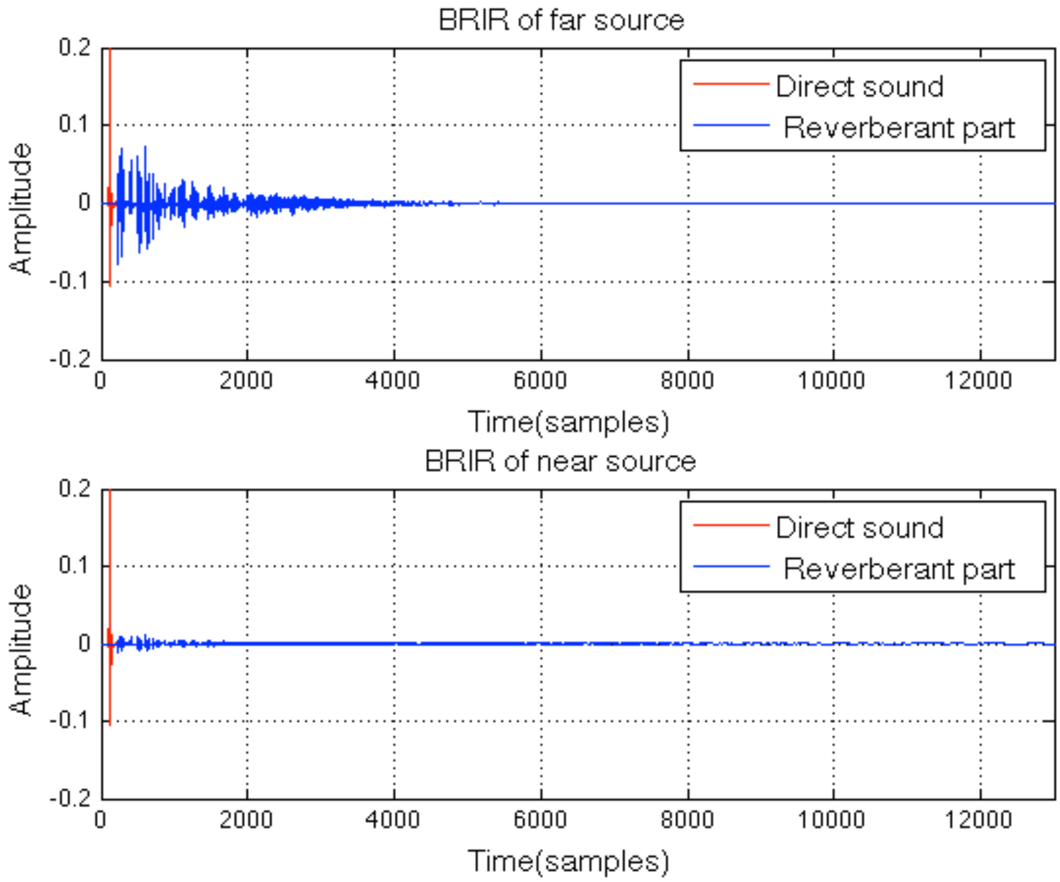


Figure 3.2: Impulse Reponses of a close and a far source. Demonstrating the D/R changes due to source location ('Close source D/R' > 'Far source D/R')

Larsen (2008) and Zahorik (2002c) investigated human sensitivity to D/R in the perception of source distance. The stimuli in Larsen's experiment were convolved with BRIRs measured with a KEMAR dummy head at a distance of 4.0m inside a room with reverberation time of 0.78 seconds. Signals were presented through headphones. The D/R was manipulated by scaling the direct path of the BRIRs (first

3ms). The levels of the sound sources were normalized across all cases, and subjects were informed to ignore small level differences and focus on the D/R discrimination. The main outcomes showed that the just noticeable differences (JDN) of D/R were 2dB to 3dB for D/R values of 0dB and 10dB and at least 5dB to 6dB for D/R values of 10dB and 20dB. Zahorik (2002c) performed similar experiments to Larsen's. However he differed in one respect: instead of scaling the direct part of the BRIRs in order to manipulate the D/R, he scaled the reverberant part. He found a constant JND of 5dB to 6dB for D/R values between 0dB and 10dB, which differed from Larsen's (2dB to 3dB). These findings are important because they show that D/R sensitivity is lower for sources with extreme D/R value (sources too far and too close respectively) and higher for sources with small D/R value (medium distances). Interestingly, Zahorik (2005) argues that the role of D/R is to provide absolute distance information rather than discriminations between relative distance changes, which are signalled with small changes in the amplitude where the human ear is very sensitive. He supports this argument of the D/R JDN he obtained in his earlier work (2002c). He further supports his argument on the large trial-to-trial variability on distance judgments observed in the listening test he performed previously (2002a) using virtual acoustics where D/R was the primary distance cue available.

The way our brain uses D/R information to localise distance is not very well understood yet. Possibly listeners are not able to separate direct and reverberant sound explicitly, so they cannot compute D/R directly (Kopčo 2011). However, there are other parameters that vary with D/R which are possibly used by the human brain to process information from the D/R. These parameters are the spectral variance, interaural coherence, interaural cross correlation (Larsen 2008) and early-to-late power ratio (Kopčo 2011).

3.1.3 Binaural cues

When a sound source is moved around the listener's head, the distance between the source and the listener's ears changes. These distance changes are perceived as interaural level differences (ILD). Level at the near ear is increased and level at the far ear decreased (Figure 3.3). The effect is stronger for closer than farther sources (Kim 2001; Shinn-Cunningham 2000b); and stronger for lateral sources and negligible for sources near the median plane. The level difference between the ears is increased dramatically across all frequencies (even at low frequencies) when the source is at a distance below 1.0m (Brungart 1999a; Kim 2001; Shinn-Cunningham 2000b). These near field ILD changes can provide absolute distance information for close sources (Kim 2001; Shinn-Cunningham 2000b). Fukuda (2003) examined the effect of ILD on relative distance perception via binaural reproduction. The virtual space was modelled with a very low reverberation time (0.1 seconds). The results indicate that the perceived relative distance of lateral angles (greater than 30°) can be altered with ILD manipulation. This shows that ILD can provide relative as well as absolute distance information in acoustic dead spaces in the near field and outside the median plane.

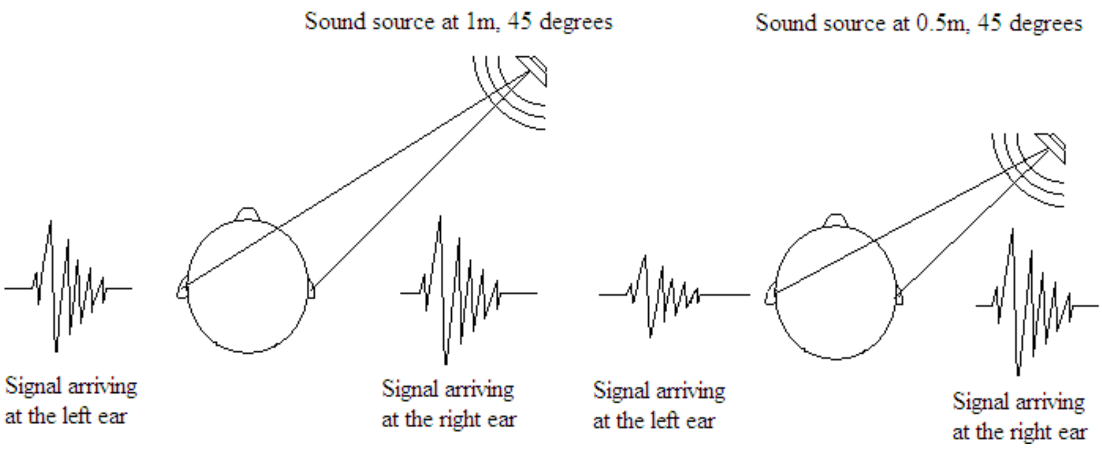


Figure 3.3: ILD increases at lateral angles when the same source is placed at distance closer to the head.

Interaural Time Differences (ITD), on the other hand, has been shown to be independent of distance changes (Brungart 1999a).

Another cue provided to the listener when the source is very close is the Acoustic Parallax (Brungart 1999; Kim 2001). Acoustic Parallax is the difference between the direction of source relative to the ear and the angle of the source relative to the centre of the head (Figure 3.4). This difference increases as the source approaches the head. Due to the parallax effect, the directional frequency response of the pinna varies with distance. When the source approaches the head, its angle relative to the ipsilateral ear is pushed away from the interaural axis, so the HRTF features of that ear are high pass filtered at increasingly lateral azimuth locations. However, this effect varies much less compared to the ILD cue that was mentioned before (Brungart 1999) and is thus considered less significant for distance perception.

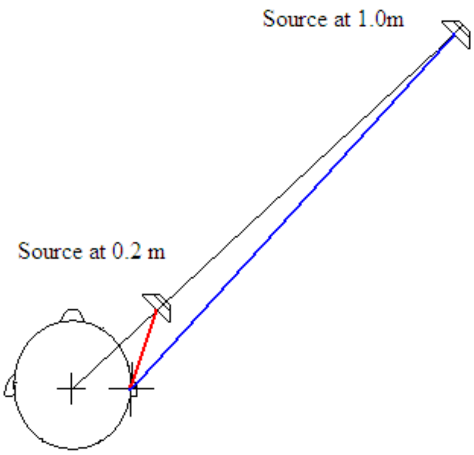


Figure 3.4: Acoustic parallax. The angle of the source relative to the centre of the head remains the same for booth distances but the angle of the farther source relative to the ipsilateral ear is increased

Reverberation has the effect of reducing the ILD at all frequencies (Ihlefeld 2004). Shinn-Cunningham (2000b) examined distance perception under reverberant and anechoic conditions for nearby sources. The sources were virtually synthesized with individual HRTFs and BRIRs. The main outcome from Shinn-Cunningham's experiment is that listeners based their judgment on reverberation and naïve listeners could not learn to use the ILD cue to localise distance even when reverberation was absent. This indicates that reverberation is a more robust cue than ILD for nearby sources.

Finally, Shinn-Cunningham (2000b, 2000c) argues that distance judgments are similarly good both under monaural and binaural conditions suggesting that binaural cues are not used for distance perception in echoic environments even for nearby sources. There is further evidence that binaural cues are not strong enough to be useful in D/R discrimination (Larsen 2008). Larsen mentions that this does not imply that distance perception is equally good via binaural and monaural listening.

3.1.4 Spectral changes due to source location

In natural environments the reflective surfaces and obstacles absorb the high frequencies more than the low frequencies. Additionally, when the source moves away from the listener, the proportion of reflected energy increases compared to the direct sound. By conflation, when the proportion of reflected energy increases (sound source is farther), the changes effected on the spectrum are more noticeable (Zahorik 2005). Also, for distances greater than 15m, low pass filtering occurs due to air absorption (Blauert 1976). Butler (1980) suggested that low pass filtered sounds are perceived farther away than high pass filtered. This difference is related to our lifetime auditory experience where air absorption decreases the high frequency spectrum of sound sources that are located at far distances. Thus, it is expected that manipulation of the high frequency spectral content of a source affects its distance perception.

Finally, it is important to note that the spectral changes discussed will be unable to provide distance localisation cues unless the listener has prior knowledge of the source characteristics (Zahorik 2005; Brungart 1998). They are thus considered as relative cues. This leads to the next topic, which relates to how familiar a certain source is to the listener.

3.1.5 Familiarity and learning

Relative distance cues (level, spectral cues) cannot provide information about the distance of the source unless the listener is familiar with the characteristics of the source. Speech is a good example, because people are very familiar with its acoustic properties which vary with the production level (an increase in the fundamental frequency and at high frequency content at high production level (shouts) (Lienard 1999). Humans perceive the location of whispered speech on average closer and shouts farther than normal speech (Zahorik 2005; Brungart 2001). Also, its has been shown that listeners are able to use these familiar characteristics of speech that vary with production level in order to make good distance judgments of live talkers Mershon (1991).

The learning factor is directly related to familiarity. Familiarity is the stored knowledge about the characteristics of sound sources (e.g. speech) and the acoustic conditions. Learning is the process during which the listeners familiarize themselves with the acoustic environment and stimulus conditions, in order to be able to localise accurately both distance and direction of the sound source.

In virtual systems learning seems to be an important non-acoustic cue and its absence degrades the performance quality of the systems (Cunningham B. 2003; Schoolmaster 2003a).

When a listener is exposed to an unfamiliar environment for the first time, without having any acoustic reference, he/she would find it very hard to judge the correct distance. However, once exposed to several different sounds with level differences, distance judgments become more accurate (Shinn-Cunningham B. 2003).

Distance perception in virtual systems is also improved with reverberation, but it can be enhanced even further with experience (Shinn-Cunningham 2000a; Devallez 2009). Schoolmaster (2003a, 2003b) performed experiments in order to investigate the effect of learning in distance perception under varying and constant room conditions. The results from his test suggest that if the room conditions are manipulated during the distance perception tests, the responses are less accurate; on the contrary, if the conditions are held constant, the responses are much more accurate. This indicates that listeners construct a "map" of distance based on recent experience (over the test session). However, in both cases (varying and constant) distance perception improved through experience.

Learning seems to play an important role in sound perception. However, investigating the effect of learning can be very time consuming. Van Wanrooij (2005) and Hofman (1998) investigated the learning time that is required for a human subject to adapt to "new ear" (new ears were their own ears with some modifications). The experiments lasted weeks, but the results were very interesting as they showed that after long term training subjects could learn to perceive sound source direction with the new ears as well as with their own ears.

3.1.7 Vision

The interaction between audio and visual has also been shown to affect auditory perception. In directional localisation, the direction of an auditory target is moved into the direction of the closer visual target for angular separations between the two of more than 30°. This effect is called *vetriloquism effect* (Gardner 1968). Similar effects have been reported in distance perception (Devalez 2009; Mershon 1980). There is also evidence that vision improves distance judgment accuracy and lowers judgment variability (Zahorik 2001).

Devalez (2008) studied the impact of visual cues in spatial impression. In a subjective test, Devalez used different types of samples which were convolved with dummy head BRIRs. The subjects were asked to adjust the D/R according to the visual cues. The D/R was manipulated by adjusting the first 2.5ms of the BRIR. Under all tested conditions, the subjects adjusted the D/R at greater values than what was originally measured.

Sound can also bias visual perception in the temporal domain. This effect is known as *temporal ventriloquism* (Morein-Zamir 2003).

3.1.8 Physical vs perceived distance

The distance of a sound source is much more difficult to be identified than its directional location. In many experiments it has been shown that there is no linear relationship between perceived distance and physical distance (Zahorik 2002a; 1997 Nielsen 1993; Mershon 1975)

According to evidence physical far distances −well beyond the critical distance, where the D/R are extreme negative values− are underestimated. It is suggested that this compression in the perceived physical distances occurs because the properties of the perceived sound signals such as IACC and spectral and temporal cues remain constant well beyond the critical distance. As a result the signal arriving at the ears of the listener is very similar to a signal of a closer source (Larsen 2008). The amount of underestimation varies among individuals (Zahorik 1997, Nielsen 1993). It has also been reported an upper limit on how far an auditory event can be perceived. This case is known as *auditory horizon*. An opposite effect exists where the distances are overestimated for physical distances very close to the listener (Zahorik 2002a, 2002d; Larsen 2008; Mershon 1975; Bronkhorst 1999; Békésy 1949). The underestimation of far physical distances and the overestimation of close physical distances is referred in the literature as *specific distance tendency*.

3.1.9 Individual HRTF and head tracking

Individualised HRTFs do not seem to affect distance perception for binaural systems, if reverberation cues are present (Zahorik 2000; Zahorik 2002b). Also, most of the research in the area of distance perception using virtual acoustics was implemented without allowing the subject to move their heads (Zahorik 2002a, 2000; Larsen 2008; Valente 2010; Kopčo 2011; Brugart 1999; Devalez 2008; Fukuda 2003). Simpson (1973) further supports that distance perception is not improved with head movements. In addition, Pelegrini (2001) argues that head-tracking is only necessary when non-individualised HRTFs are used. As such, it appears that there is a trade-off between the use of individual HRTFs against providing subjects with the ability of moving their heads in facilitating correct distance perception.

3.2 Externalisation

A common problem with binaural reproduction systems is the poor sound externalisation. A sound source is externalised when it is perceived outside the radius of the head (Blauert 1976; Begault 2000). If one source is more externalised than the other, it would mean that the virtual auditory event would be perceived farther from the head radius than the other. Thus, distance perception in binaural systems is intrinsically linked to the ability of externalising sound.

3.2.1 Early stages of inside head locatedness (IHL) research

Blauert (1974) defined the problem of IHL in terms of perceived distance of a virtual sound source. In his book he reviews experiments and hypotheses around inside head locatedness. Some of the hypotheses about the causes of this problem were: the loading of the eardrum with impedance different from that of a free sound field; lack of head movements; the similarity of the signals in the two ears; lack of the pinna effect; absence of the sound energy presented at the rest of the body besides ears; differences in the transfer characteristics between electrostatic transmission channels; overmodulation of the nervous system; and natural resonances of the headphones.

Green (1988) further suggested that human listeners have ‘pushbuttons’ in their head that detect the presence of headphones and automatically perceive images inside their heads.

3.2.2 Individual cues

Weinrich (1992) designed and tested a system that improves externalisation in a headphone stereophonic system using a cross-feed delay network. He concluded that the frequency range between 4kHz and 12kHz is essential for externalisation. Consequently by introducing individual cues in this frequency range, better externalisation can be achieved.

Individualised HRTFs seem to improve externalisation. Kim (2005) compared the performance of different HRTFs in binaural reproduction. The results for his subjective tests suggest that both non-individualised and individualised HRTFs can provide externalisation but the individualised one to a higher degree. This finding is also supported by Volk (2008). Although some research has been conducted, the effect of non-individualised HRTFs in externalisation is not yet well understood.

The amount of influence of individualised HRTFs on externalisation is highly dependent on the source signal. Begault (2000) investigated the effect of individualised HRTFs on the externalisation of a virtual speech source. The results from the subjective test showed that individualised HRTFs do not improve the localisation accuracy, externalisation and reversal rates. This outcome is also supported by Møller’s (1996) research, which states that individualised HRTFs do not improve localisation accuracy of speech signals. This probably occurs because the spectral energy of speech exists in the frequencies where ITD cue is dominant and pinna cues are very weak (Begault 2000).

3.2.3 Decorrelation of ear signals

Decorrelation of left and right ear signals is a very important factor that affects externalisation. Decorrelation in real life occurs due to the frequency dependent diffraction of the head, early reflections and reverberation (Kendall 1995). In the median plane, where the signals in both ears are very similar, decorrelation can be provided via head movements. Kendall (1995) described the design of decorrelation filters and implemented subjective tests to observe the effect of decorrelation in source externalisation. Some of the advantages of decorrelation filters were described as:

1. Elimination of colouring effects
2. Production of diffuse sound fields.
3. Production of enough externalisation in headphone playback.
4. Elimination of image shift.
5. Removal of Precedence effect.

Brookes T. (2005) studied the effect of left and right ear asymmetry in sound externalisation. This asymmetry decorrelates even further the signals arriving at left and right ears. Brookes used dummy head recordings made in a 250m^2 concert hall for distances between 1.0m and 10m, and at 0^0 azimuth and elevation. The source signal was a 10 seconds long male speech and some percussive instruments. He made the recordings both with symmetric and non-symmetric pinnae. The subjective tests showed that asymmetric pinnae recordings provide significantly greater externalisation over symmetric pinnae recordings. Consequently he proposed slight modification of one of the HATS's pinnae before performing generic HRTFs or BRIRs measurements in order to produce this desirable asymmetry.

3.2.4 Head Tracking

Wightman (1999) examined if real-time updating of the listeners HRTFs according to his/her head position via head tracking improves externalisation. In his model he used individual anechoic binaural recordings of white Gaussian noise and individual headphone equalisation. His results suggest that head-tracking does not play any role in the externalisation. On the other hand, Loomis (1990) supports that head-tracking can provide externalisation even with the pinna cue absent. Wersenyi (2008) performed subjective test with HRTFs of a good localiser and the result of the subjective test showed an improvement of approximately 30% on externalisation with the head tracking on.

3.2.5 Reverberation

Addition of reverberation improves externalisation and realism in headphone simulation significantly (Begault 2000, 1992; Durlach 1992; Shinn-Cunningham B. 2000a).

Völk F. (2009) examined the effect of the BRIR length in sound externalisation. He used a measured BRIR of a good localiser. The results suggest that high order reflections up to 25ms increase the perceived externalisation but reverb time over 100ms does not provide extra externalisation.

Finally the value of D/R, which is directly related to reverberation, also seems to be important in externalisation (Sakamoto 1975)

3.2.6 Headphone equalisation

The quality of binaural reproduction can be increased if the headphones are equalised correctly (Hammershøi 2005; Griesinger 2008). The equalisation depends mostly on the coupling between the individual's ears and the headphones. Most of the open type headphones, as the ones that were used in this research (Sennheiser HD800), work as a volume cavity system for frequencies below 4kHz (Figure 3.5). Above this frequency standing waves are built (Masiero 2011). As a result, the pressure at the eardrum changes depending on the headphone fitting. Therefore, many important spectral cues can be affected, and the quality of the binaural reproduction is degraded.

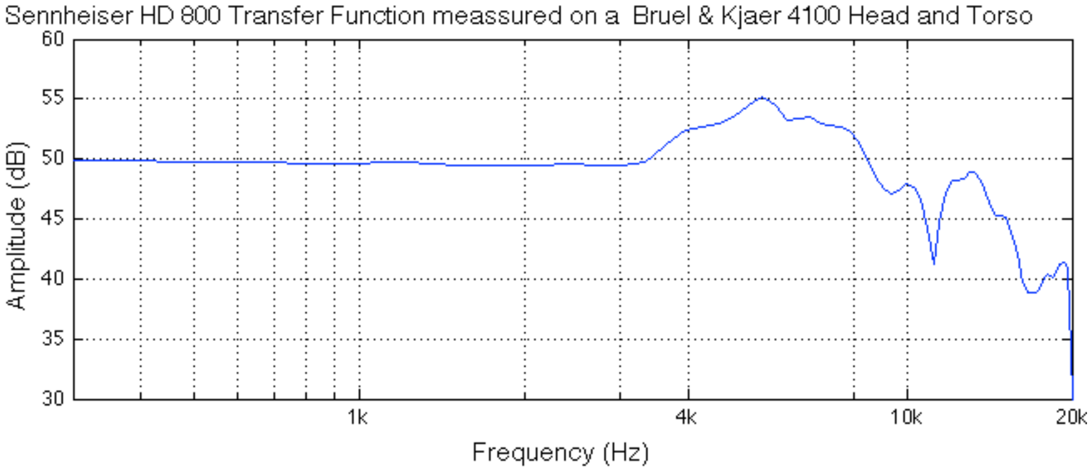


Figure 3.5: Headphone Transfer Function of HD 800 measured with B&K 4100

In order to equalise the headphones, the output signal of the headphones has to be convolved with the inverse Headphone Transfer Function (HpTF).

The HpTF can be measured on a HATS or on an individual by inserting miniature microphones inside its ears. The HATS HpTF can produce artifacts because it will differ significantly from the individual's HpTF. Variations in HpTF can lead to unwanted effects that can influence externalisation and localisation performance. Therefore, single HpTF is not a good solution (Hammershøi 2005; Griesinger D. 2008). Additionally the spectral differences on the individual HpTF due to the headphone fitting can create noticeable effects (Paquier 2010). In order to overcome this problem, technicians suggest measuring individual HpTF for different headphone fitting positions and then averaging them. Massiero (2011) proposed a robust headphone equalisation method that can minimize most of the problems of the previous equalisation methods.

Kim (2005) designed equalisation filters using the wiener filter approach (Oppenheim 2010). He conducted four sets of test: (1) KEMAR HRTFs (Gardner 1994) without headphone equalisation; (2) KEMAR HRTFs with headphone equalisation; (3) Individual HRTFs without equalisation; (4) Individual HRTFs with equalisation. The subjects were informed about the location of the sound source because the determination of angle and distance can be very stressful, especially when the source is localised inside the head. His main outcomes are: (1) that individual headphone equalisation is crucial for externalisation and that individual HRTFs are important for consistent distance perception. This result contrasts with Zahorik's (2003) paper whose listening test results support that individualised HRTFs are unimportant for correct distance localisation (2) lateral angles are easier to externalise, which is also supported by Völk (2008).

Good externalisation without headphone equalisation has been achieved previously in distance perception tests under reverberant conditions (Zahorik 1997; Kopčo 2011).

3.2.7 Virtual vs Real source

Some researchers have defined externalisation, as a measure of how indistinguishable is a virtual from a real source. Hartman (1996) and Langendijk (2000) performed direct comparisons of virtual sources presented through headphones and real sources reproduced through loudspeakers. The drawback with this method is the interference of the sound wave generated from the loudspeaker and the earpad. A novel method to eliminate this effect has been proposed by Moore (2008). Moore used a DSP technique in order to make the headphones acoustically transparent so that the signal reaching the listener from the loudspeaker would sound as if the headphones were absent. The analysis of the results showed that headphone transparification works well for fixed listener position but the high frequencies are attenuated. Further work is needed in order to minimize this effect.

3.3 Conclusion

In this section the literature review related to the project is addressed. It covered distance perception and the cues that affect it, which is the main subject of this research project, as well as the topic of sound externalisation.

I) For distance perception the most important cues are:

Source level: Is a relative distance cue unless the listener has prior experience with the level of the source.

D/R: The ratio between direct and reverberant energy is a strong absolute distance cue.

Binaural cues: ILD and acoustic parallax can provide absolute distance information for close distances (below 1.0m). However, when reverberation is present, subjects tend to rely on the D/R instead of the binaural cues.

Signal spectrum: Low frequency sounds are perceived farther away because they are associated with low frequency air absorption. This is a relative distance cue unless the listener has previous knowledge about the source spectrum.

Familiarity: Familiarity with the characteristics of the source can facilitate humans make better distance judgments.

Learning: After having been exposed to the new listening conditions for some time, listeners calibrate their perception to these new conditions and are able to make better distance judgments.

Vision: Improves accuracy and reduces the variability of the responses.

Individualised HRTFs, head tracking and headphone equalisation do not seem to be very important factors for distance perception under binaural reproduction through headphones, especially when the virtual environment is reverberant.

II) Externalisation of binaural reproduced sound sources can be improved with reverberation, decorrelation of left and right ear signal, headphone equalisation and head tracking. The effect of individualised HRTFs on externalisation is not yet fully understood. Better externalisation occurs at lateral angles.

4. Modelling and Measuring the BRIR

In this section the methods used for modelling and measuring BRIRs will be demonstrated.

4.1 Modelling of BRIR

The BRIR anatomy was described in section 2.4. The three parts of the response (Direct sound, early reflections, late reverberation) were modelled separately. Furthermore, the sound field in each ear was computed individually taking into account their different positions in space.

4.1.1 Design of the direct part

The KEMAR diffuse field equalised HRTF database was used for the BRIRs design (Gardner 1994). The interaural amplitude of the HRTF was corrected using Supper's (2010) method as follows. The HRTF database was measured at distance of 1.4m from the centre of head. When the loudspeaker moves around the listener, the distance between each ear and the loudspeaker changes, which results in a change of sound pressure level at each ear. Should this not be corrected, a distance related ILD cue (see subsection 3.1.3) for a source at 1.4m would remain in the response. Hence, this ILD was removed in order to re-introduce later the correct ILD depending on the new source location. The correction was implemented by calculating the distance between the loudspeaker and each of the ears for every measurement position of the HRTF database. Then according to the source position, the level at the near ear was decreased and the level at the far ear increased. The effective diameter of the head for this correction was assumed to be 21cm. This resulted in a maximum change in ILD (at 90^0 azimuth) of 1.3dB.

The ILD cue for every given source location can then be approximated by adjusting the signal level on the left and right ear correspondingly (Shinn-Cunningham 2000d), (Equation 4.1). This was applied in every modelled sound source by determining the ratio between the distance of right ear and virtual source (d_{right}) over the distance between the virtual source and the centre of the head (d_{centre}) and the ratio between the distance of the left ear (d_{left}) and the virtual source over the (d_{centre}). These ratios were then multiplied by the right and left ear Head Related Impulse Responses respectively in order to model the distance related ILD. Changes were applied across all frequencies equally.

$$Signal_{right} = \frac{d_{right}}{d_{centre}} \times HRIR_{right}$$

$$Signal_{right} = \frac{d_{left}}{d_{centre}} \times HRIR_{left}$$

Equation 4.1: Direct sound modelling

Ideally, distance dependent HRTFs should have been used, as specific features of the HRTF related to distance change are neglected using the above correction (Spors 2011). However, the method used is considered a sufficient compromise since these features do not change dramatically for distances above 1.0m (Brungart 1999) and the nearest source modelled for this project is at 1.10m.

4.1.2 Design of the early reflections

A geometrical room acoustic technique called Image Source Method was used for the design of the early reflections. Following this technique, reflections are modelled as ideal specular reflections, and the positions of the reflections are obtained by mirroring the original sound over each wall surface of the room (Allen 1979) (Figure 4.1). Reflection up to any order can be modelled by mirroring any image source produced on the walls or obstacles of the room.

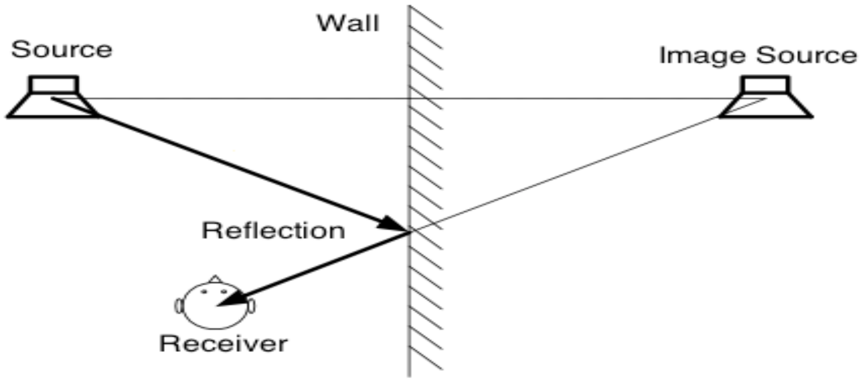


Figure 4.1: Image source example

The method is adequate for wavelengths that are very small compared to the room surfaces (waves that "see" the room surfaces as infinite plates) as the behaviour of sound waves are modelled following the principle of light rays. Sound pressure is obtained by considering energy rays rather than complex pressure waves (Allen 1979; Kapralos B. 2006). Diffusion was ignored and reflections were modelled as ideal specular reflections. Diffraction effects and wave interference are also not considered by this method. As such, the image source method is not the most advanced method for reflections modelling. Other geometric techniques such as Ray Tracing, Beam Tracing or combinations of these methods can provide better predictions (Vorländer 2007). Despite these shortcomings, the method is here considered sufficient to represent sources in a room with no obstacles and at frequencies of interest in a simple and computational efficient way. The model used for this project was programmed in Matlab.

Reflections arriving up to 80ms after the direct sound were modelled using this method. Reflections beyond 80ms, and with reflection orders greater than one, are typically assumed to arrive equally from all directions and can be described as exponentially decaying noise (Kapralos B. 2006) as will be described in subsection 4.1.3. The incoming direction for each modelled early reflection was determined in order to render it using the nearest available HRTF angle from the KEMAR database. This resulted in a maximum resolution error of 2.5˚ in the horizontal plane and 15˚ in the median plane. This model could have been improved via HRTF interpolation. However, unlike the direction of the direct sound source, there is no proof in the literature that the direction of reflections affects the perception of distance. This is also supported by the outcome of this research project.

An average absorption coefficient was used for all the walls of the modelled rooms and this absorption was the same for all the frequencies. The absorption values were taken from average published data (Cox 2009). Equation 4.2 demonstrates how the obtained image sources were processed.

$$Image_L = \frac{r_L}{I_L} \times HRIR_L \times \left(\sqrt{1-a}\right)^{order}$$

$$Image_R = \frac{r_R}{I_R} \times HRIR_R \times \left(\sqrt{1-a}\right)^{order}$$

Equation 4.2: Processing of image source

Where:

$Image_L$:	*Absorption coefficient of the room walls*
$Image_R$:	*Order to the image source (how many times it has been reflected on the room boundaries)*
r_R:	*Left ear HRIR of the image source location*
r_L:	*Right ear HRIR of the image source location*
I_R:	*Distance between image source and left ear*
I_L:	*Distance between image source and right ear*
$HRIR_R$:	*Distance between source and left ear*
$HRIR_L$:	*Distance between source and right ear*
a:	*The image source arriving at the right ear*
order:	*The image source arriving at the left ear*

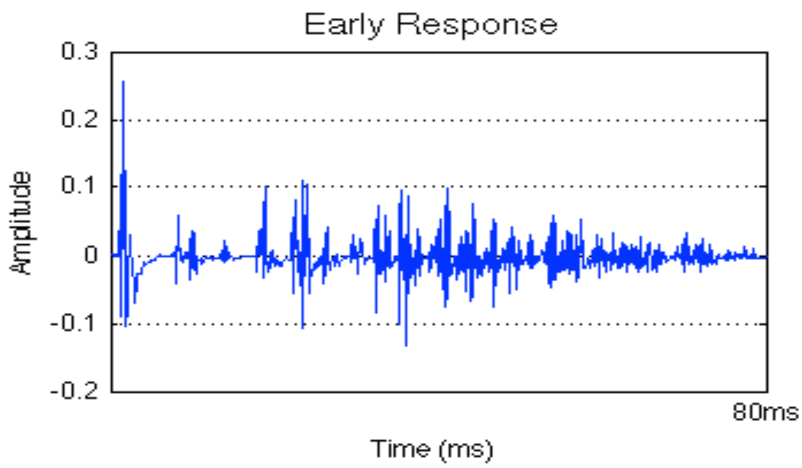


Figure 4.2: Modelled early response

4.1.3 Modelling reverberation

Menzer (2010) argues that late reverberation of BRIR can be modelled accurately by considering the time-frequency energy decay relief and frequency interaural coherence. In this BRIR model only the first factor was considered. The late reverberation was modelled using uncorrelated Gaussian noise for each ear, which was shaped using separate decaying functions for each octave band within the range 63Hz to 8kHz. The formula for designing the decaying function of the octave band was taken from Zahorik (2009) (Equation 4.3)

$$d(t) = 10^{-3\frac{t}{T_{60}}}$$

Equation 4.3: Octave band decay function

Where:
t: time
T_{60}: reverberation time of the octave band

The reverberation time of the octave bands can be estimated using Sabine's reverberation equation with published absorption coefficients at different octave bands or by estimating the reverberation time at each octave band via measured impulse responses.

Figure 4.3 provides an example of the modelled reverberation tail of each octave band. These octave bands are then added together in order to form the reverberation tail.

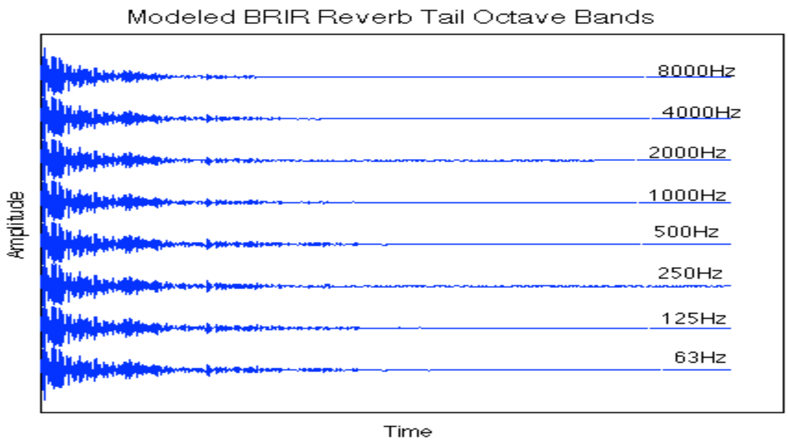
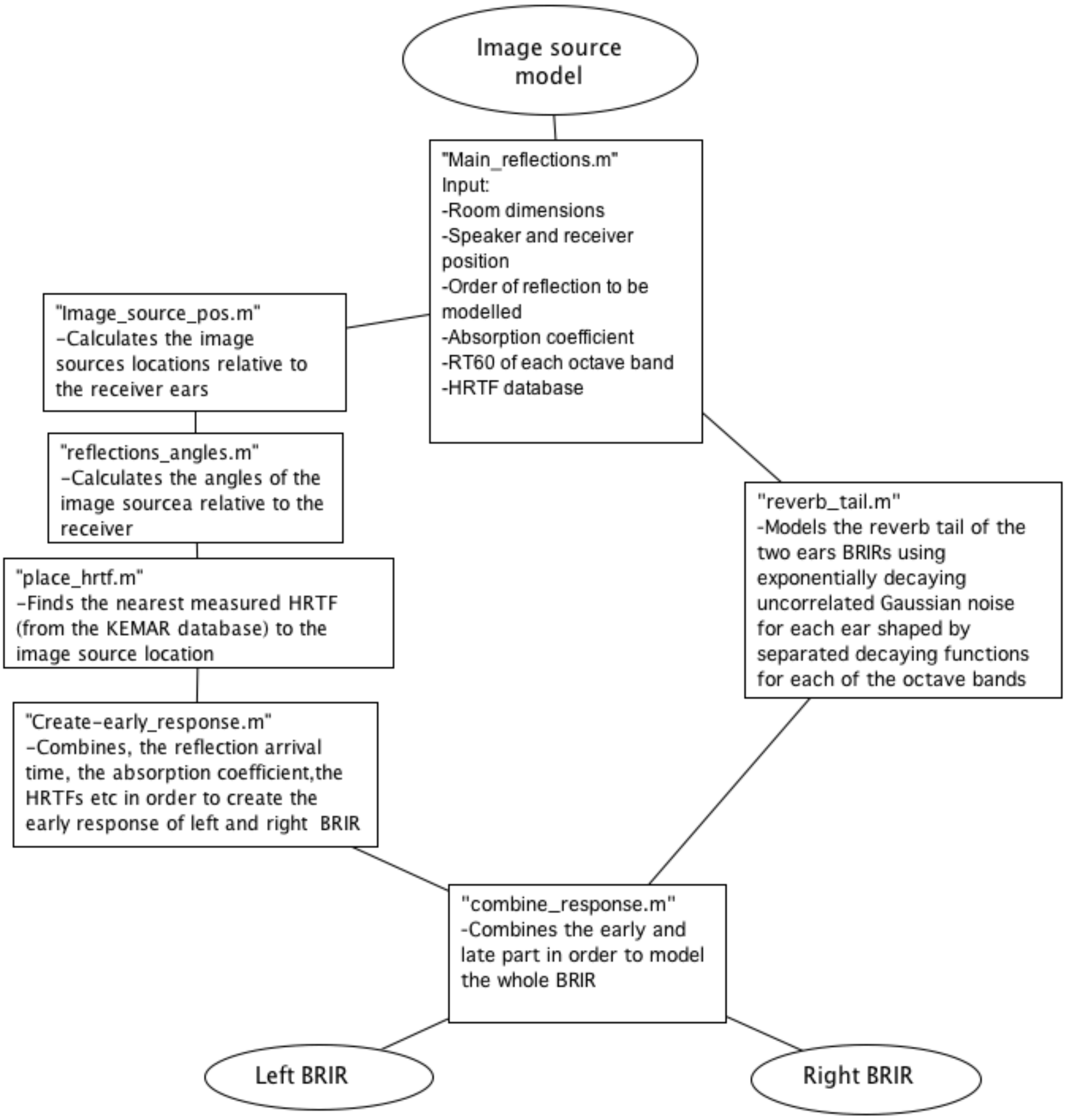


Figure 4.3: BRIR reverb tail octave bands

4.1.4 Combining Early Reflections and Reverberation

The method of combining early reflections and late reverberation was also taken from Zahorik (2009). The combination was performed by matching the RMS level of the last 15ms of the early response of length T_e (= 80ms) with the first 15ms of the broadband reverberation tail. Then the first $T_e - 15$ms of the late response were removed and finally the responses were added together (Figure 2.3).

4.1.5 Code block diagram



4.2 Measurement of BRIR

Generic BRIRs were measured. These BRIRs were used to evaluate how well modelled BRIRs described in section 4.1 provide relative distance information in comparison to real BRIRs.

The measurements took place in the listening room at the University of Salford. The dimensions of the room are 6.6m × 5.8m × 2.8m, the background noise level is 5.7dBA, and the reverberation time is 0.27ms. This room is full of diffusers and meets the requirements of ITU-R BS 1116-1 for subjective assessments of small impairments in audio systems

The HATS was placed 5.0m away from the front wall, in the middle of the sidewalls and 1.30m above the floor (Figure 4.4). It was attached to a chair which was placed on a turntable with angles printed on.

The speaker used for the measurements was a Genelec 1030A active near field monitor which was driven for all measurements with a fixed power level of 80dB SPL at 0.5m. Three measurements were performed at distances of 1.10m, 1.70m and 2.62m; all at 45° azimuth.

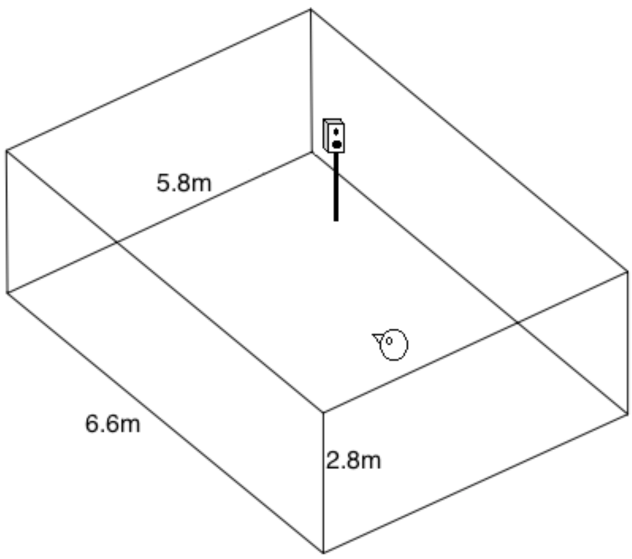


Figure 4.4: Receiver and speaker inside the listening room

The HATS used for the measurements was a B&K 4128D which has fixed microphones in the blocked ear canals. The outputs of the microphones were connected to a Norsonic front-end type 336 preamp. The outputs of the preamp were sent into Fostex VC-8 DA/AD converter, which in turn were sent to PC-based Adobe Audition 2.0.

BRIRs were extracted with Aurora 4.2, a plug-in for Adobe Audition. The excitation signal for each measurement was a 15th order MLS (Vanderkooy 1994; Farina 2000) repeated 40 times in order to improve signal to noise ratio. D/A and A/D conversion was effected with 16-bit precision at a sampling frequency of 44.1 kHz.

DC components were present in the measurements but were removed using a tool of Aurora 4.2 which is called "remove DC components". This tool actually applies a high-pass filter on the sound file. Finally, the BRIRs were truncated in Matlab using a 13000 samples long rectangular window, resulting in responses 0.29 seconds long.

4.3 Match modelled BRIRs with Measured BRIRs

Modelled BRIRs were designed to match the BRIRs measured in the listening room (section 4.2) in order to test their perceptual differences in providing relative distance cues.

The first order reflections were modelled using the method described in 4.1.2 for a rectangular room with dimensions same as the listening room where the BRIRs were measured. The broadband absorption coefficient of the modelled listening room's walls was estimated from the mean reverberation time of the measured BRIRs octave bands.

The octave band analysis of the measured BRIRs was implemented using a Matlab function called 'fdesign.octave.m' and a butterworth filter using 'filter.m' function. The reverberation time of the octave bands was estimated from the energy decay curves.

Thus, the octave bands reverberation times of the measured BRIRs were used in Equation 4.3 to achieve similar energy decay relief between modelled and measured BRIRs.

The octave analysis worked well up to 8kHz and down to 63Hz, but outside this range the Butterworth filter used for the analysis became unstable. Therefore, octave bands below 63Hz and above 8kHz were excluded. Hence, modelled BRIRs have narrower frequency range.

The energy decay curves of the modelled and real listening room BRIRs differ slightly (graphs A.5, A.6 in Appendix A). Also, their frequency responses differ considerably for frequencies above 8kHz (Figure A.3, A.4). The reasons for these differences are: (1) The broadband absorption coefficient used for the first 80ms of the modelled BRIRs, which was constant across frequency and identical for all room boundaries; (2) All the reflections of the modelled BRIRs were ideal specular reflections which included no diffusion effects; (3) The measured BRIRs were measured with a different HATS, not the KEMAR which was used to model the early reflections and direct sound.

Finally, the near ear and far ear D/R of the modelled and measured BRIRs are not equal (Figure 4.5). However, the maximum difference in the near ear between modelled and measured is 1.4dB. This is smaller than the just noticeable difference, and in the far ear the difference is 2.1dB which is slighter higher than the JND (2dB). However, the far ear D/R is not a strong distance cue (Larsen 2008; Zahorik 2002c).

This can affect slightly the relative distance perception between sources modelled with real and measured binaural impulse responses. For example, for sources synthesized with real BRIRs, the D/R difference between sources at 1.10m and 1.70m is 2.5dB; but when the same sources are synthesized with modelled BRIRs, the D/R difference is 3.6dB. This may be perceived as a slightly larger relative distance in the case where sources are synthesized with modelled BRIRs.

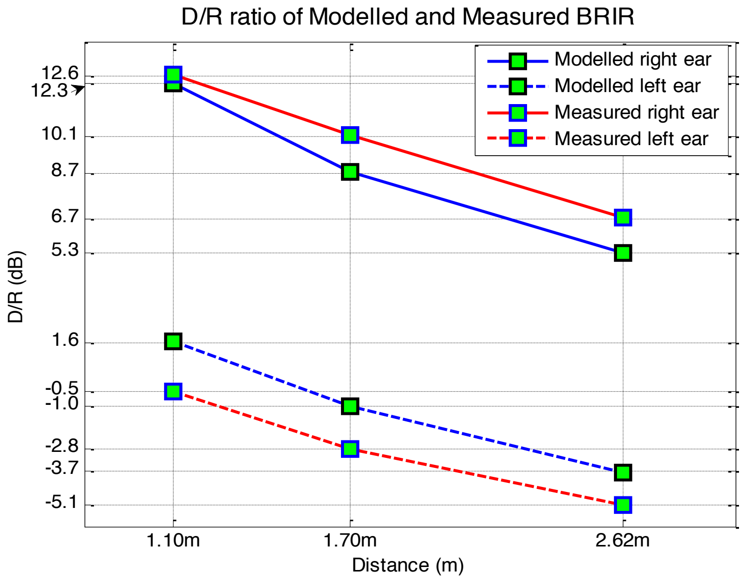


Figure 4.5: Direct-to-reverberant ratio of modelled and measured BRIRs for both near and far ear. The D/R was determined by performing time windowing in the BRIRs in order to separate the direct and the reverberant part. The direct part was selected as the first 3ms of the BRIR and the reverberant part as the remainder. Finally, the power ratio between them was calculated.

4.4 Conclusion

In this section, the methods used for modelling and measuring BRIRs were discussed. The modelled BRIRs have some limitations in comparison to real ones. These are: 1) Reflections are idealised as specular, with no diffusion or diffraction effects; 2) All the walls have been modelled with the same broadband absorption coefficient; 3) Reverberation includes octave bands only up to 8kHz; 4) No interaural coherence matching was performed between measured and modelled BRIRs; 5) HRTFs were not fully distance dependent (only the ILD was manipulated).

5. Listening Tests Methodology

The section starts by reviewing typical methods used for testing auditory distance perception. Later the methodology used for the design and running of the subjective tests together with the design of the test stimuli and the experimental set up are discussed.

5.1 Review on relative and absolute distance judgments

Absolute distance judgment is referred to as the actual egocentric perceived distance of the sound source. For example, '*source 1'* is perceived at 5.0m. The 'relative' term, on the other hand, means that there is some reference external to the listener that he/she can use to compare. Relative distance judgment between two sources may be:

- Categorical: Identifying which sound is closer or farther and not how much closer or farther the sounds are. Therefore, the answer to such a listening test would be binary, 'far' or 'close', and no information regarding the absolute distance of the sources or the relative distance between them would be given.

- Continuous: Judging how much closer or farther the sound sources are. This distance judgment can be either relative or absolute (relative judgment of egocentric distance). For example, if '*sound 1'* was 2.0m away from the listener and '*sound 2'* 3.0m away from the listener and the subject indicates that '*sound 1'* is at 1.0m and *'sound 2'* at 2.0m, he/she would be inaccurate in an absolute sense but accurate in a relative sense.

Many different procedures have been used for reporting perceived absolute distance. These methods include direct marking of the sound source distance in explicit scales (metres or feet) by verbal report or by typing it using a numeric keypad (Zahorik 2002a; Kopcko 2011; Pellegrini 2002) and (Nielsen 1993; Mershon 1975). Some researchers have used implicit scales such as the Thurstonian and other continuous scales (Zahorik 1997; Picinali L. 2010). All methods give reliable results but some of them require more testing time (Zahorik 1997).

For continuous relative distance testing (how much closer or farther sounds are) Devalez (2008) used the MUSHRA method (Bech 2006) in order to perform direct comparison of all the sounds within a set. The subjects had to rate the relative distance of 8 different virtual sources at varying distances between 1.0m to 8.0m against a reference sound, on a scale ranging from 0 to 10 dimensionless units through steps of 0.1 units. Devalez suggests that this method can prevent the bias from learning effect and high variability of responses across trials that would occur if common evaluation methods such as the A/B comparison were used. Fukude (2003) (binaural reproduced sound sources) and Strybel (1984) (loudspeaker sources) conducted pairwise comparisons and asked subjects to judge which source was farther or closer. Zahorik (1997), carried out pairwise comparison of binaural reproduced sound sources using Thurstonian scaling. The main task was to indicate which sound of the pair appeared egocentrically closer.

5.2 Experimental Definitions

The experiment reported here was separated into three parts. In all parts, accuracy of relative distance perception (continuous relative distance judgment) was investigated. Absolute distance of sound source was not part of the experiment.

Three logarithmically spaced distances in feet (1 foot= 30.48cm), 1.10m, 1.70m and 2.62m were chosen to be included in the relative distance perception test. Distances were chosen to be logarithmic spaced as it has been previously used by Zahorik in several experiments (2002a, 2002c, 1997). He has also estimated a power function which is equivalent to a linear function that can relate logarithmically transformed perceived distances to physical distances with an average slope of approximately 0.4 (2002a, 2002c).

Listening test Part 1 objectives

1) To investigate accuracy in relative distance perception between two virtual sound sources in the absence of level differences between the sources.

2) To evaluate perceptual differences between modelled and measured BRIRs in providing relative distance cues.

Listening test Part 2 objectives

1) To investigate the effect of direct-to-reverberant ratio on distance perception.

2) To examine the effect of early reflection TOA on distance perception.

Listening test Part 3 objectives

Investigate relative distance perception with level differences between the virtual sound sources.

5.3 Design of the test stimuli

5.3.1 Source signal

A 4 second Italian speech sentence, presented at 45° azimuth and 0° elevation, was selected as the source signal. Speech is a familiar sound to humans and has been previously used for distance perception tests via binaural reproduction. Distance simulation and externalisation of speech sources even without headphone equalisation and individual HRTFs can be very satisfactory (Zahorik 2002a; Völk 2009; Begault 1992; 2000; Kim 2005; Valente 2010; Devalez 2008;Philbleck 2002). In addition, using non-personalised HRTFs in binaural reproduction when speech signal is processed does not affect the localisation accuracy; most of the spectral energy of speech is in those frequencies where ITD is a significant strong cue and less energy is at the frequencies related to the pinna cues, which are person specific, and important for front-back, externalisation, and elevation localisation (Begault 2000; Møller 1996)

(Figure 5.1). The source was always presented at 45° because lateral sources are externalised easier compared to frontal sources, especially when artificial HRTFs are used (Völk 2008; Kim 2005). The results from this experiment might be less applicable to other types of source signals and angular locations because it is argued that information from each distance cue is processed in different ways depending on source direction and source signal (Zahorik 2002a). The source was chosen to be in the horizontal plane because ILD of lateral angles provides an extra absolute and relative distance cue. Also, externalisation is greater for sources outside the median plane (Volkl 2008; Kim 2005). Finally, the speech sample was selected to be in Italian, in order to ensure that subjects, mostly English speakers, would not focus on the semantic context but rather on the acoustic effects of the presented speech.

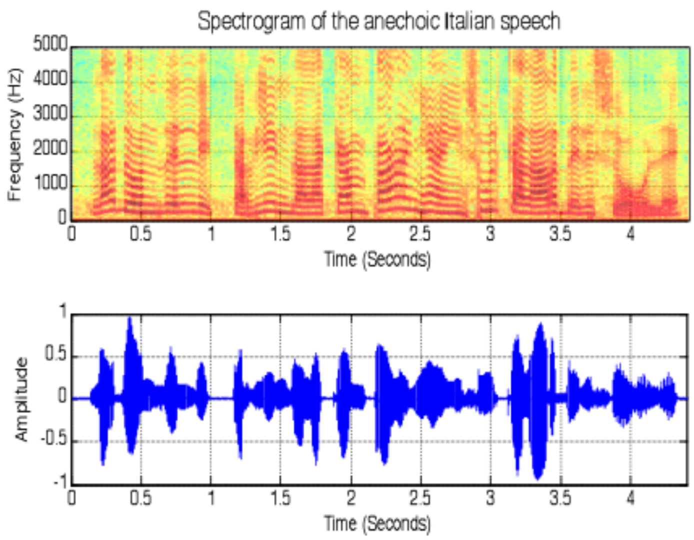


Figure 5.1: Spectrogram of anechoic Italian speech

5.3.2 Listening Tests: Part 1

In order to follow the objective of listening test Part 1 (section 5.2), the level differences between the sounds presented at different distances ideally should be zero. By convolving the BRIRs, level disparities between the right (near) ear signals of the sounds were created. The reasons for this is that modelled and measured BRIRs are normalised by their peak value (all of them have maximum value of 1) and are not RMS normalised. As is clearly noticeable in A and C graphs in Figure 5.4, the reverberant part of the BRIRs at 2.62m has more energy than the reverberant part of the BRIRs at 1.10m. So by convolving the anechoic signal with the BRIR at 2.62m more energy (greater amplitude) will be added to the signal than when it is convolved with the BRIR for source at 1.10m. These amplitude differences had to be removed (Figure 5.2). The responses were thus normalised using the mean RMS level for the right ear signals. All the mean RMS amplitudes were matched to the lowest mean RMS amplitude of these signals. In order to keep the distance related interaural level differences unaffected, all the stimuli both left and right signals were multiplied by the same ratio. This ratio is the minimum mean RMS amplitude of all the right ear signals over the mean RMS amplitude of the right ear signal of each virtual source (see Equation 5.1 and Figure 5.3). Natural differences on the left ear signals, which occur due to the distance related ILD (Figure 3.3), were thus maintained.

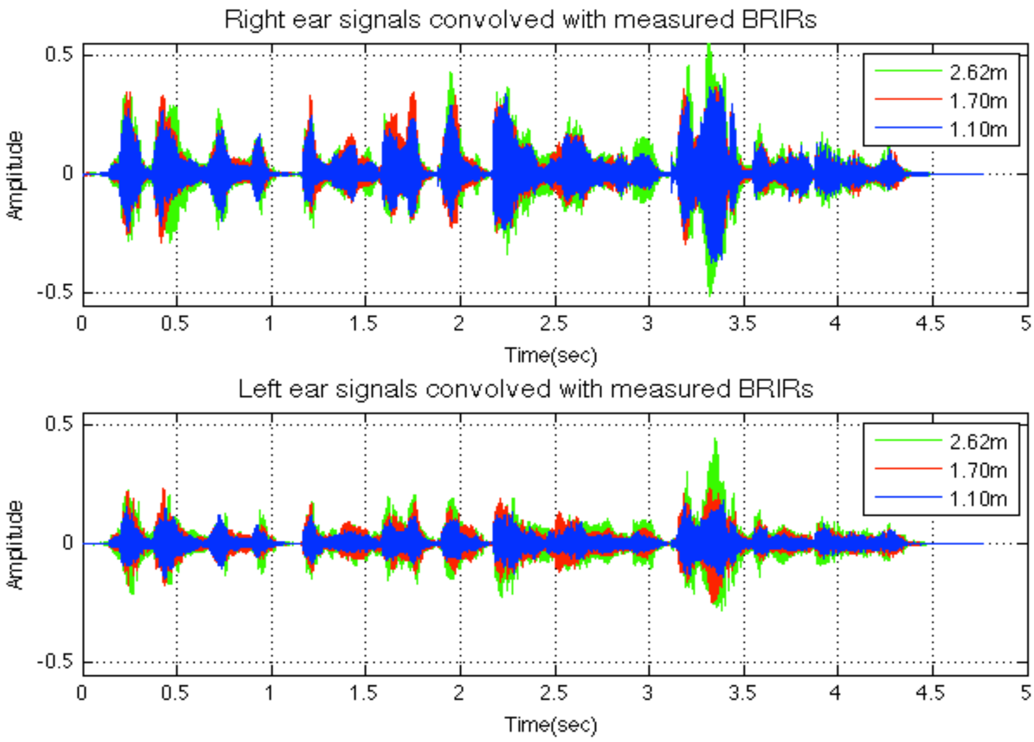


Figure 5.2: Right (top) and left (bottom) ear signals convolved with measured BRIRs without RMS normalization

$$Norm\_Signal_{right} = Signal_{Right} \times \frac{RMS\_Right_{min}}{RMS\_Right}$$

$$Norm\_Signal_{left} = Signal_{left} \times \frac{RMS\_Right_{min}}{RMS\_Right}$$

Equation 5.1: RMS Normalization

Where:

RMS_Right_{min}:	*minimum RMS all right ear signals*
RMS_Right:	*RMS of the right ear*
$Signal_{right}$:	*right ear signal before RMS matching*
$Signal_{left}$:	*left ear signal before RMS matching*
$Norm\_Signal_{right}$:	*right ear signal after RMS amplitude normalization*
$Norm\_Signal_{left}$:	*left ear signal after right ear RMS amplitude normalization*

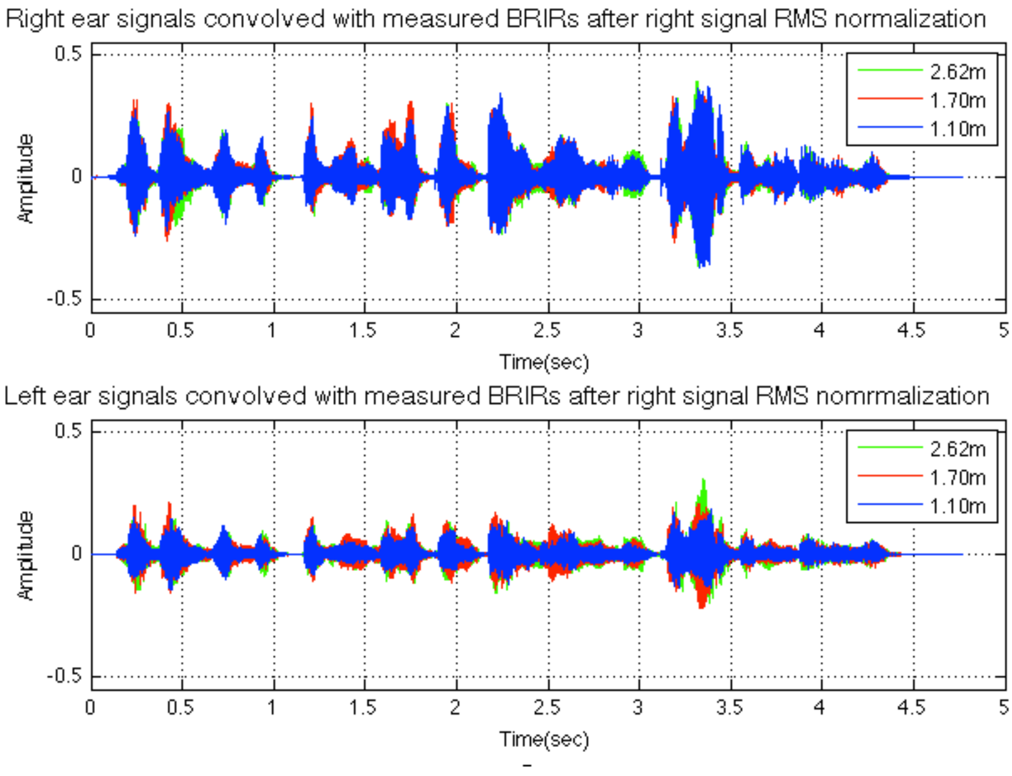


Figure 5.3: Right (top) and left (bottom) ear signals convolved with measured BRIRs after RMS matching

The fact that the mean level of the right ear signals was normalized does not mean that the level of signals at the right ear is going to be the same over the whole duration. For example at 1.2 seconds the signal of the virtual source at 1.70m is stronger than the signal from the virtual source at 2.62m, but at 3 seconds is the other way round (Figure 5.3). This occurs because the reflections TOA of the signals are different. The normalization removed the monaural level cue at the near ear; but it did not so for the monaural level cue on the far ear due to the distance law (Level at the far ear for source at 2.62m is approximately 0.8dB greater than at 1.10m). Previously researchers have performed a technique called 'level roving' to eliminate far ear level cue in order to make the subject focus on cues other than amplitude in distance perception (Kopčo 2011; Brungart 1999c) and it worked for almost all subjects apart from a few exceptions. However, these two researchers looked into distances below 1.0m where differences between far signals are up to 10dB. This technique was not applied here since the test samples all relate to distances larger than 1.0m where differences in the farther ear are, as seen above, below 1dB.

5.3.3 Listening Test: Part 2

For this experiment only modelled BRIRs were used. In order to evaluate how well the manipulation of D/R can simulate distance and to investigate the effect of the reflection TOA on distance perception, six new BRIRs were created. All examples below involve an underlying BRIR for a given distance where either the D/R or the reflection TOA were manipulated to match that of a source in another distance.

- BRIR at 1.10m with D/R of source at 1.70m
- BRIR at 1.10m with D/R ratio of source at 2.62m
- BRIR at 2.62m with D/R ratio of source at 1.70m
- BRIR at 2.62m with D/R ratio of source at 1.10m
- BRIR at 1.10m with reflections TOA of source at 2.62m
- BRIR at 2.62m with reflections TOA of source at 1.10m

The D/R ratio of the BRIRs was defined by manipulating the levels of the early reflections and reverberation. The delays between reflections and direct time, directional characteristics of the reflections, and ILD due to distance changes remained unchanged for the first 4 BRIRs listed above. For example BRIR at 1.10m with D/R ratio of source at 2.62m was modelled by replacing the levels of the reflections (early and late) of the BRIRs at 1.10m with the levels of the reflections of the modelled BRIRs at 2.62m (Figure 5.4).

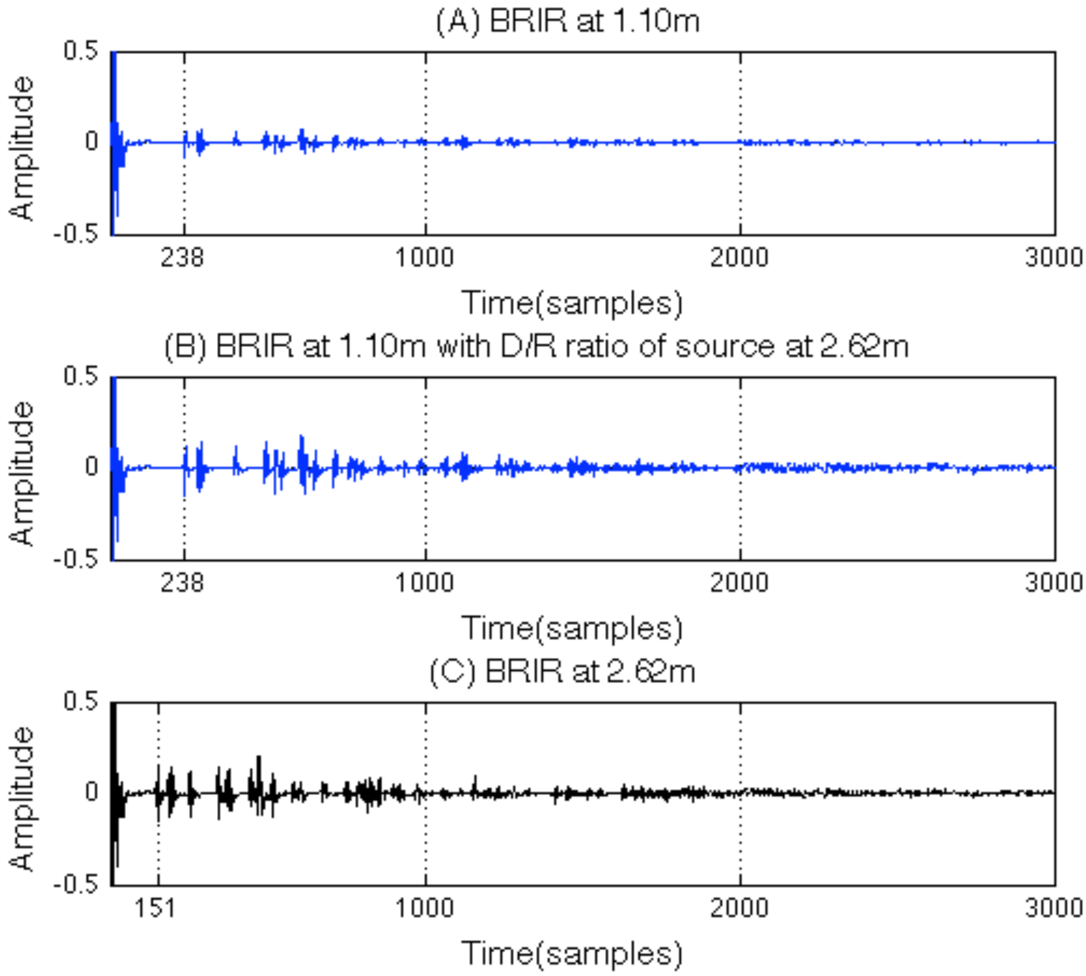


Figure 5.4: D/R ratio manipulation vs real source displacement

Figure 5.4 (A) and (B) shows the same BRIRs but with different D/R. The D/R of (B) is the same as (C) but with different reflection TOA and directional characteristics. Reflection TOA and the ILD of (B) are the same as (A).

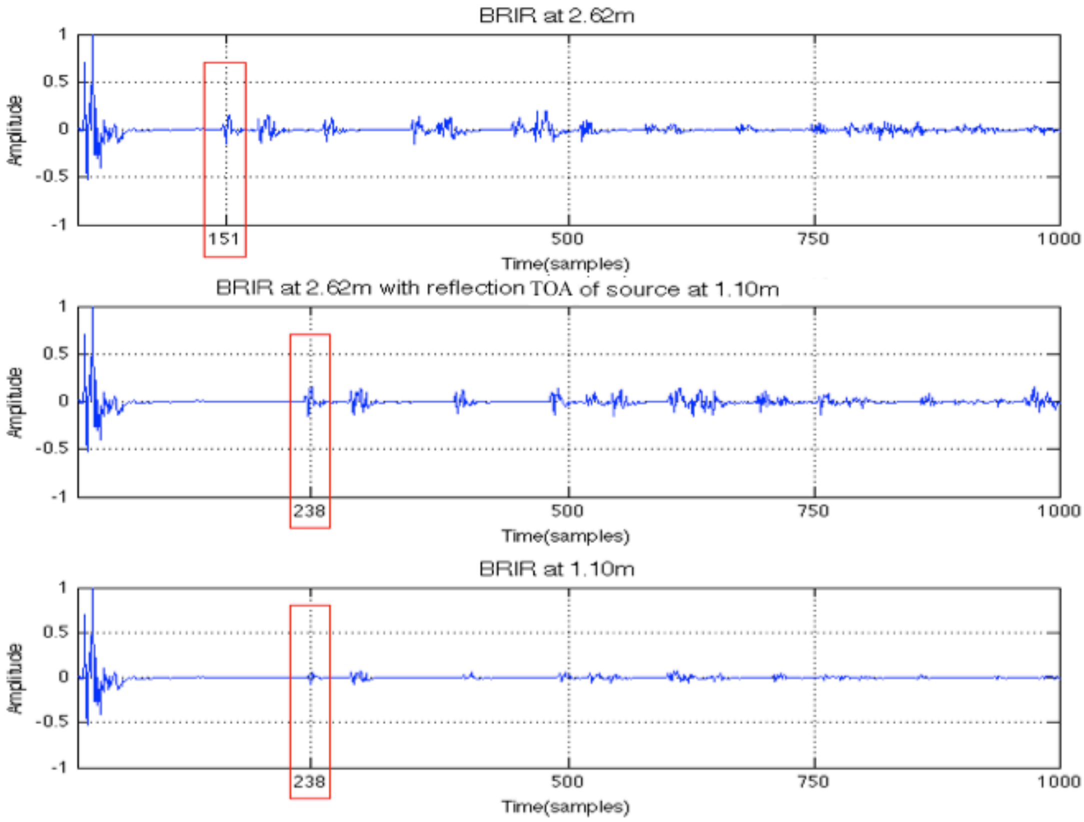


Figure 5.5: Delay manipulation. The reflections from the first two BRIRs have the same level and directional characteristics but different TOA. The reflections on the last two BRIRs have the same delay times but different levels and directional characteristics.

BRIR at 2.62m with reflections TOA of source at 1.10m was created by replacing the reflections TOA of the modelled BRIRs at 2.62m with reflections TOA of the BRIR at 1.10m (First of all, the reflection TOA of the source at 1.10m were stored in array. Then, the code in Matlab was edited in way that instead of using the reflection TOA of source at 2.62m to model the BRIR to use the reflection TOA of the source at 1.10m). Therefore, both BRIRs have the same D/R, directional characteristics, and ILD, but their reflections TOA differs (Figure 5.5).

Finally, the level normalization was implemented exactly as in the previous part (subsection 5.3.2) in order to remove the distance cue arising from level differences.

5.3.5 Listening Test: Part 3

This part of the experiment is identical to that described in Part 1 but this time all samples include their natural level cues. The level of the three virtual sources convolved both with modelled and measured BRIRs was manipulated by multiplying the anechoic speech signal with the distance ratio of the reference source which was the 1.10m, over the desirable distance (Figure 5.6).

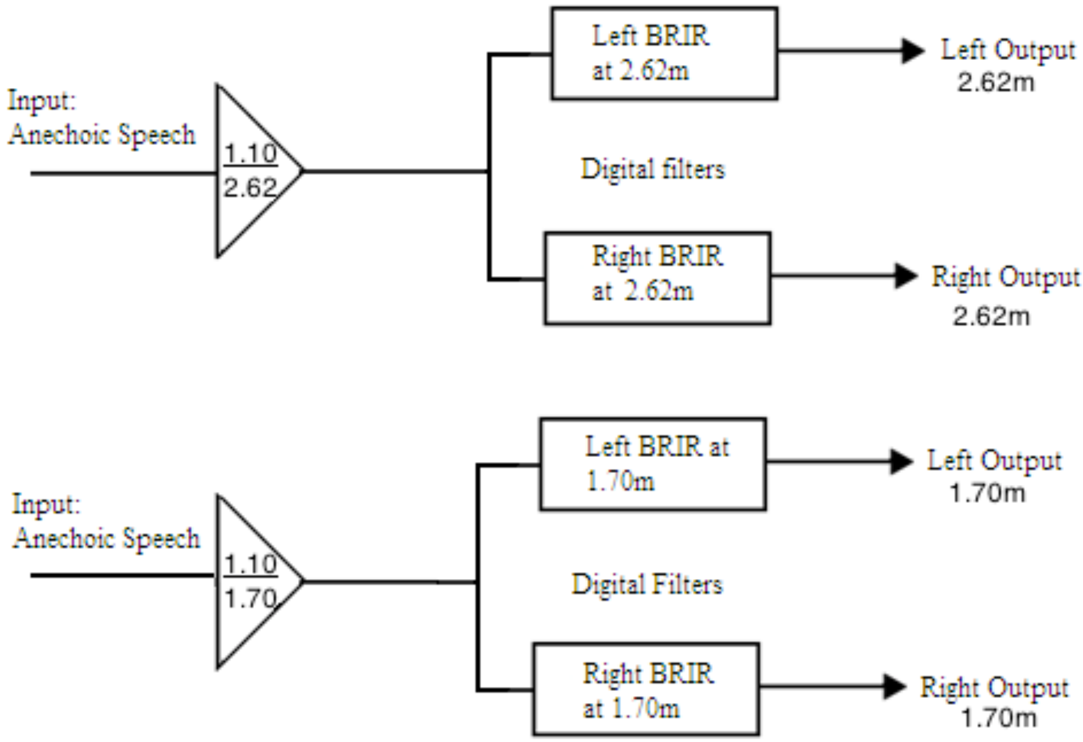


Figure 5.6: Level manipulation

5.4 Listening Test Methodology

5.4.1 Subjects

Twenty-three unpaid listeners participated in the test (20 male and 3 female, aged from 20 to 42; 32 years old average). Twelve were postgraduate acoustics students, three undergraduate acoustics students, and eight staff members of the acoustics department. Most of them have participated in listening tests more than 10 times, but none of them in any tests related to distance perception.

5.4.2 Procedure

The whole test was implemented without any breaks between the three parts, and the overall duration was approximately 25 minutes. The test took place at the University of Salford semi-anechoic chamber with a background noise of 3.8dBA. Sounds were generated via a Matlab script at sample frequency 44.1kHz and presented via Sennheiser HD 800 open-circum-aural dynamic stereo headphones. Sound levels remained constant in the range around 70dBA for broadband stimuli.

Part 1 was always implemented first and Part 3 last. The stimuli pairs in every part were randomized. Before starting the test, subjects were given written (appendix B) and oral instructions about the test process. Before starting Part 1, they listened to all the six stimuli included in this part randomly as many times as they wanted in order to evaluate the context of the test. The same was done before Part 3 because a new cue, the level differences, was introduced.

Subjects gave their answers on a computer user interface (Figure 5.8). They were presented with pairs of sounds with 0.2 second silence gap and they were asked to judge how farther or closer they perceive the second sound compared to the first in a dimensionless continuous scale ranging from -0.5 (closer) to + 0.5 (farther) with a middle point of 0 (equal) and with a resolution of 0.05 steps. Therefore, relative distance accuracy was measured in dimensionless units and not in metres or in feet (Figure 5.7).

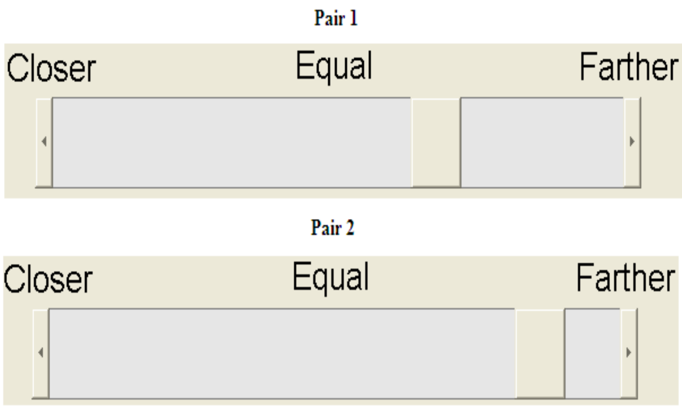


Figure 5.7: Example of scaling relative distance. Relative distance between second sound and first sound in Pair 2 is larger than in Pair 1.

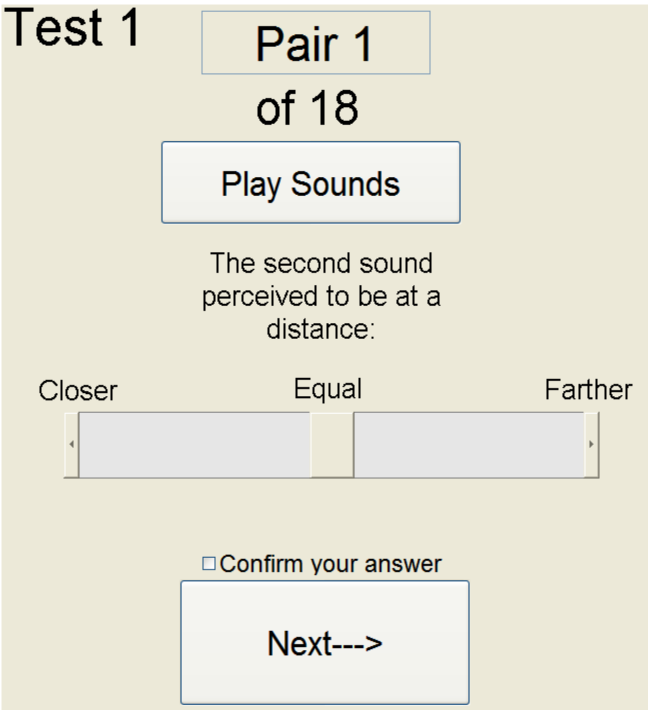


Figure 5.8: User interface for relative distance evaluation.

For each listening sample there was a button to start the playback of the pair of samples ‘Play sounds’, a checkbox to confirm the choice ‘confirm your answer’ and a next button ‘Next→’ to move to the next pair. The subjects could repeat the pairs as many times as they wanted by pressing the ‘Play’ button (Figure 5.8).

5.4.3 Organization of test samples

In order to make the analysis in the next section easier, abbreviations for the names of the test files were used (Table 5.1)

Full sentences	Abbreviations
Distance 1.10m	1
Distance 1.70m	2
Distance 2.62m	3
Convolved with modelled BRIR at distance X (without level cue)	modX
Convolved with measured BRIR at distance X (without level cue)	mesX
Convolved with modelled BRIR at distance X with D/R ratio of BRIR at distance Z	modX_DRZ
Convolved with modelled BRIR at distance X with reflection delays of BRIR at distance Z	modX_delayZ
Convolved with modelled BRIR at distance X (with level cue)	modX_Level
Convolved with measured BRIR at distance X (with level cue)	mesX_Level

Table 5.1: List of abbreviations

Examples of Abbreviations:

- "Speech convolved with modelled BRIRs at 2.62" is written as "mod3" and " Speech convolved with measured BRIRs at 1.70m" as "mes2" (see subsection 5.3.3)
- "Speech convolved with modelled BRIR at 2.62 whose D/R was replaced with the D/R ratio of the modelled BRIR at 1.10m" is refereed as "mod3_DR1" (see subsection 5.3.4)
- "Speech convolved with modelled BRIR at 2.62m whose reflections TOA was replaced with the reflections TOA of the modelled BRIR at 1.10m" is written as "mod3_delay1" (see subsection 5.3.4)
- "Speech convolved with modelled BRIRs at 2.62 and level cue active" was written as "mod3_Level"

Listening test Part 1 pairs:

Pairs	Sound 1	vs	Sound2
Pair1	mod1	vs	mod2
Pair 2	mod1	vs	mod3
Pair 3	mod2	vs	mod3
Pair 4	mes1	vs	mes2
Pair 5	mes1	vs	mes3
Pair 6	mes2	vs	mes3

Table 5.2: Comparison pairs of listening tests Part 1

Listening test Part 2 test pairs:

Pairs	Sound 1	vs	Sound2
Pair1	mod1	vs	mod1_DR2
Pair 2	mod1	vs	mod1_DR3
Pair 3	mod3_DR1	vs	mod3
Pair 4	mod2_DR2	vs	mod3
Pair 5	mod1	vs	mod1_delay3
Pair 6	mod3	vs	mod3_delay1

Table 5.3: Comparison pairs of listening tests Part 2

Listening test Part 3 test pairs:

Pairs	Sound 1	vs	Sound2
Pair1	mod1_Level	vs	mod2_Level
Pair 2	mod1_Level	vs	mod3_Level
Pair 3	mod2_Level	vs	mod3_Level
Pair 4	mes1_Level	vs	mes2_Level
Pair 5	mes1_Level	vs	mes3_Level
Pair 6	mes2_Level	vs	mes3_Level

Table 5.4: Comparison pairs of listening tests Part 3

Important notes:

1. All the pairs in every Part were repeated three times each. Therefore, the whole test included 54 comparison pairs.
2. Part 1 was implemented first, then Part 2 and finally Part 3.
3. All the pairs (including their repetitions) in every part of the test were randomized.
4. Sounds convolved with modelled and measured BRIRs were not compared directly due to the small differences in the D/R ratio and the limitations of the modelled BRIRs that could possibly make the comparison confusing (section 4.3).
5. Level of the near ear signals in Part 1 and Part 2 were normalized.
6. Headphone equalisation was not applied.

5.5 Conclusion

In this section the test set up and the methodology of the listening test were described together with the design of the test stimuli. The section also includes the list of abbreviations made for shortening the names of the test samples in order to make the data presentation and the analysis better.

6. Results analysis

In this section the results and the statistical analysis are presented.

The threshold of consistency of each the subject's responses was set to 67% and only subjects with consistency above this limit were included in the analysis. This agreement was used within each subject as a measure or performance. Consistency in between the subject's responses is defined as how consistent are the judgments of each subject in terms of which source is closer or farther, and is not related on whether the response is right or wrong. For example, if on the first part a subject had given 15 wrong answers out of 18, which is a consistency of 83%, would be included in the analysis because this would mean that he/she perceived consistently sources wrongly farther or closer, so he/she did not judged randomly. The limit was chosen to be at 67% which mean that subjects were allowed to make 6 inconsistent judgments out of the total 18 of the first listening test, which translates to one inconsistent answer per three repetitions of each pair (in part 1 they were 6 pairs and each of them was repeated 3 times).

From the analysis five subjects were excluded because they had very low performance level (the consistency between their responses in part 1 was between 53% to 61%), so is very likely that these subjects were judging randomly. All these five subjects were all male. A possible reason for this low performance level could possibly be that the subjects did not pay attention during the test. For example, one subject was found asleep during the test process. The subjects included in the analysis had a minimum consistency of 83% and ten of them a consistency of 100%, so their performance level was very high.

6.1 Distance perception of synthesized sources with normalized level

The estimated relative distances are presented in a Box-and-Whisker plot, showing the lower and upper quartile values, and the median value. The whiskers represent the remainder of the data, with extreme values displayed as a red cross. Positive values in the plot indicate that *sound 2* was perceived farther than *sound 1* and negative values indicate that *sound 2* was perceived closer than *sound 1*. The three distances tested 1.10m, 1.70m, and 2.62m correspond to the abbreviations 1, 2, and 3 respectively.

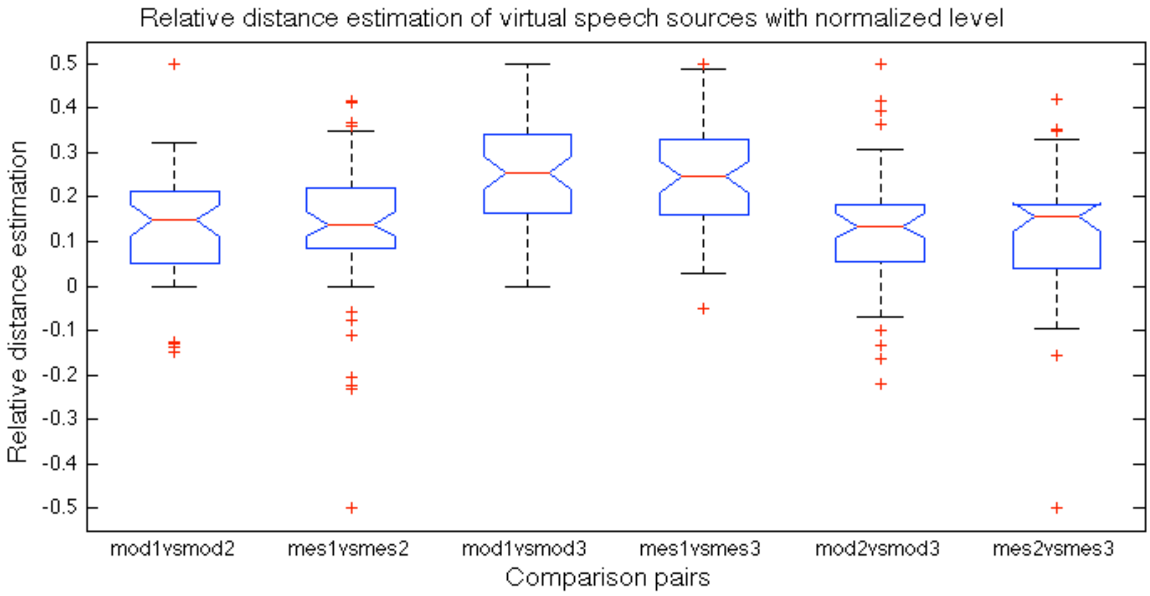


Figure 6.1: Box-and Whisker-plot showing the estimated relative distances between virtual sources with normalized level.

Second sound perceived:	mod1vsmod2	mod1vsmod3	mod2vsmod3	mes1vsmes2	mes1vsmes3	mes2vssmes3
Farther	85.20%	98.15%	79.60%	85.20%	98.15%	79.60%
Same distance	5.60%	1.85%	9.40%	1.80%	0.00%	11.00%
Closer	9.20%	0%	11%	13.00%	1.85%	9.40%

Table 6.1: Percentage cases where the second virtual source was perceived farther, closer or at the same distance from the first source (virtual sources with normalized level).

Table 6.1 and Figure 6.1 suggest that subjects could indeed detect which source is farther or closer without level cues for all the comparison pairs both for modelled and measured BRIRs.

The next aspect that is evaluated is the reliability of the subjects' to perceive small or large relative distance. For example, whether mod3 is perceived much farther from mod1 in comparison to mod2. For this evaluation, a Kruskal–Wallis one-way analysis of variance (ANOVA) was implemented. Kruskal–Wallis is a non-parametric ANOVA and it was used because the data were not normally distributed (Hollander 1999).

Relative distance between	vs	Relative distance between	Level of Significance
Mod1vsMod2	vs	Mod1vsMod3	p<0.0001
Mod2vsMod3	vs	Mod1vsMod3	p<0.0001
Mes1vsMes2	vs	Mes1vsMes3	p<0.0001
Mes2vsMes3	vs	Mes1vsMes3	p<0.0001

Table 6.2: Level of significance between different relative distances of sources with normalized level.

The results from Table 6.2 show that the relative distance differences between the pairs are significant. The results suggest that subjects can reliably discriminate short (1.10m to 1.70m) from the longer (1.10m to 2.62m) distances/

Another Kruskal–Wallis ANOVA was performed to identify whether there are significant differences in the results obtained for modelled and measured sources.

Relative distance between	vs	Relative distance between	Level of Significance
Mod1vsMod2	vs	Mes1vsMes2	p=0.88
Mod1vsMod3	vs	Mes1vsMes3	p=0.36
Mod2vsMod3	vs	Mes2vsMes3	p=0.96

Table 6.3: Level of significance of relative distances between modelled and measured virtual sources with normalized levels.

Results on Table 6.3 suggest that there are no significant differences in the results obtained either with modelled or measured responses. This result is encouraging as it suggests that the limitations of the modelled BRIRs (see section 4) do not appear to affect the accuracy of relative distance perception in these tests.

6.2 Effect of D/R and reflections TOA

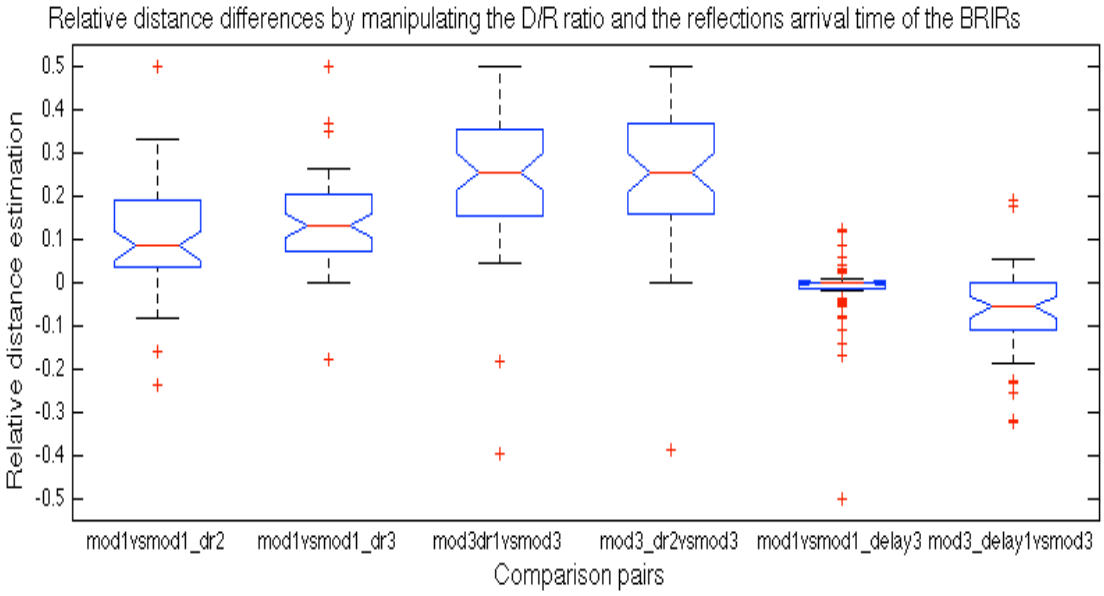


Figure 6.2: Box-and Whisker-plot displaying the estimated relative distances between modelled virtual sources whose D/R was manipulated (first four boxes from the left) and sources whose reflection TOA was manipulated (last two boxes on the right). Level cues have been normalized for all cases.

Second sound perceived:	mod1 vs mod1_Dr2	mod1 vs mod1_Dr3	mod3_Dr1 vs mod3	mod3_Dr2 vs mod3	mod1 vs mod1_delay3	mod3_delay1 vs mod3
Farther	83.30%	96.30%	94.4%	90.74%	20.37%	7.41%
Same distance	9.30%	0%	3.56%	7.4%	50%	29.59%
Closer	7.40%	3.70%	2.04%	1.86%	29.63%	63%

Table 6.4: Percentage cases where the second source was perceived farther, closer or equal to the first source when D/R or reflections TOA were the only factors manipulated from the BRIRs.

The first four boxes on Figure 6.2 and the first four columns on Table 6.4 suggest that D/R is an effective cue in providing the effect of distance perception. Manipulation of D/R provides distance cues even when all other cues in the response are kept constant. Conversely, manipulation of reflection TOA does not appear to provide reliable cues for distance perception.

Another Kruskal–Wallis ANOVA was performed in order to evaluate if significant differences in relative distance perception cues can be manipulated by changing the D/R.

Relative distance between	vs	Relative distance between	Level of Significance
mod1 vs mod1_dr2	vs	mod1vs mod1_dr3	p<0.0001
mod3_dr2 vs mod3	vs	mod_Dr1vs mod3	p<0.0001

Table 6.5: Level of significance of relative distances between source whose D/R was manipulated.

Table 6.5 indicates that the sole manipulation of D/R is indeed effective at providing larger or smaller relative distance cues. For example, manipulating the response of source at distance '1' with the D/R of source at distance '3' places it further away than modulating it with the D/R of source at distance '2', as expected. Finally, a Kruskal–Wallis ANOVA was performed (Table 6.6) to evaluate if D/R manipulation produces the same relative distance perception as a real displacement of the source. For example, can we simulate the inherent relative distance change that exists between the responses of sources at 1.10m and 1.70m by simply applying the D/R conditions of a source at 1.70m to the response of 1.10m? The statistical non-significance of the results shows that we can for all cases tested.

Relative distance between	vs	Relative distance between	Level of Significance
mod1 vs mod2	vs	mod1 vs mod1_Dr2	p=0.18
mod1 vs mod3	vs	mod1 vs mod1_Dr3	p=0.75
mod2 vs mod3	vs	mod3_Dr2 vs mod3	p=0.52
mod1 vs mod3	vs	mod3_Dr1 vs mod3	p=0.99

Table 6.6: Level of significance of relative distances between Part 1 and Part 2 pairs of the listening test.

6.3 Distance perception of synthesized sources with level differences

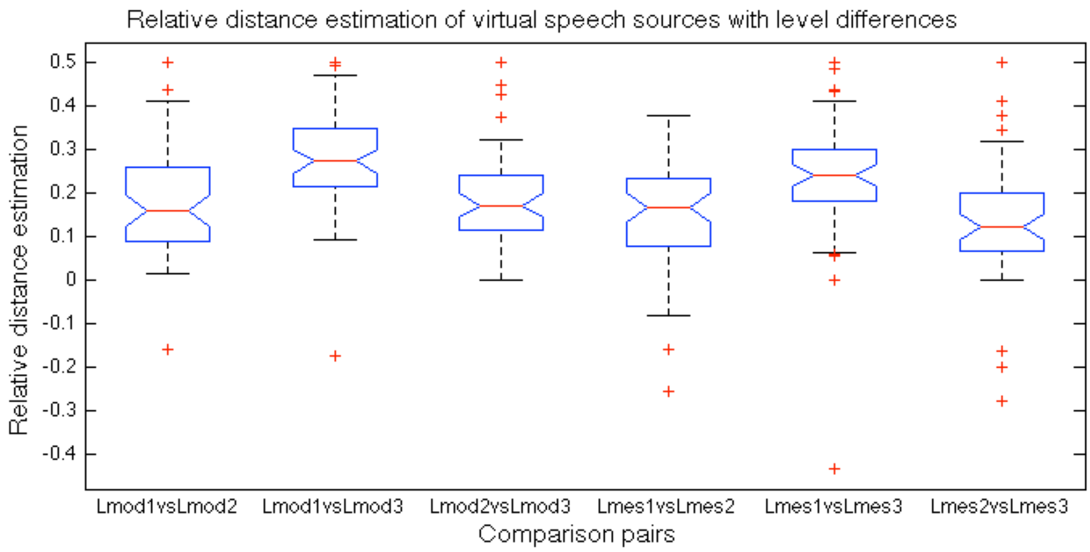


Figure 6.3: Relative distance estimation of virtual sources with level differences.

Second sound perceived:	mod1vsmod2	mod1vsmod3	mod2vsmod3	mes1vsmes2	mes1vsmes3	mes2vssmes3
Farther	98.15%	98.15%	96.3%	85.2%	94.45%	87%
Equal	0%	0%	3.7%	9.3%	3.7%	7.4%
Closer	1.85%	1.85%	0	5.5%	1.85%	5.6%

Table 6.7: Percentage cases where the second virtual source was perceived farther, closer or equal from the first source (virtual sources with level differences between them).

In order to evaluate how reliably people can judge if relative distance difference is small or large when level cue is active another Kruskal–Wallis ANOVA was conducted.

Relative distance between	vs	Relative distance between	Level of Significance
LMod1vsLMod2	vs	LMod1vsLMod3	p<0.0001
LMod2vsLMod3	vs	LMod1vsLMod3	p<0.0001
LMes1vsLMes2	vs	LMes1vsLMes3	p<0.0001
LMes2vsLMes3	vs	LMes1vsLMes3	p<0.0001

Table 6.8: Level of significance of relative distances between sources with level differences.

The Figure 6.3 and Tables 6.7 and 6.8 show that listeners can estimate accurately small or large changes in relative distance when level differences exist between them. It can also be observed that the relative distance error (which source is closer or farther) is very small.

Next, two more Kruskal–Wallis ANOVA were conducted in order to see if there is any significant difference in the relative distance perception judgments between pairs with normalized levels and pairs with level differences (Table 6.9). For example, do subjects perceive similar relative distance between virtual sources ranging from 1.10m to 1.70m when level is normalized as when it is not? The statistical non-significance of the results indicates that subjects perceive the same relative distance in both cases for almost all pairs.

Relative distance between	vs	Relative distance between	Level of Significance
Mod1vsMod2	vs	LMod1vsLMod2	0.12
Mod1vsMod3	vs	LMod1vsLMod3	0.19
Mod2vsMod3	vs	LMod2vsLMod3	**0.02**
Mes1vsMes2	vs	LMes1vsLMes2	0.64
Mes1vsMes3	vs	LMes1vsLMes3	0.78
Mes2vsMes3	vs	LMes2vsLMes3	0.61

Table 6.9: Level of significance between the same virtual source with and without level differences

6.4 Conclusion

In this chapter the results were presented and analysed. The main outcomes from this analysis are discussed in the next chapter.

7. Discussion

Modelled and measured BRIR produced similar relative distance results. Therefore, the limitations of the BRIR design did not affect the relative distance judgments. However, this does not mean that measured and modelled BRIRs would give the same absolute distance judgment. There was not direct comparison between modelled and measured BRIRs due to their differences in spectrum and D/R ratio that would affect distance judgments (section 4).

Contrary to Zahorik's finding (subsection 3.1.2), the results from the work presented here suggest that D/R is indeed a strong relative distance cue. Subjects in this listening test were able to detect relative distance accurately with normalized level between sources whose D/R difference was 2.6dB, which is less than the D/R JND obtained from Zahorik (2002c). Therefore, the detectable difference in D/R is smaller than the JND obtained from Zahorik (5dB to 6dB for sources with D/R values between 0 and 20dB). This outcome is also supported by a more recent research on D/R JND conducted by Larsen (2008), who argues that D/R JND is 2dB to 3dB for sources whose D/R is between 0dB and 10dB (The D/R of the virtual sources in this experiment was between the range 12.6dB and 5.3dB which is very close to the range 0dB and 10dB where Larsen identified the 2dB to 3dB JND (Figure 4.5). Since the strength of D/R as a relative distance cue depends on D/R sensitivity, the outcome that D/R is a strong relative distance cue cannot be expanded to sources much farther away because the D/R sensitivity is decreased.

The relative distance judgments between pairs with normalized level and pairs with level differences were similar. However, the relative distance error (which source is closer or farther) was slightly smaller (≈10%) between pairs with level differences between the sounds (Table 6.2 and Table 6.7). The reduction of relative distance error was expected with the addition of level differences because it has been confirmed previously that sources with lower levels are perceived farther away than sources with higher level (Mershon 1975; Garden 1968).

The manipulation of reflections TOA does not seem to provide any distance information. Additionally, the D/R on its own can provide the same relative distance changes as when all the factors of the BRIR related to a change in source distance are adjusted accordingly. Therefore, subjects do not base the relative distance judgment on the other factors of the BRIRs related to a change in the source distance −the reflections TOA, the direction of the reflections, and the ILD. Shinn-Cunningham (2000b) also argues that listeners base their judgments on reverberation and not on ILD, even for distances closer than the ones tested here where the ILD is a much stronger cue.

The information above shows that to move a source at different distances in a virtual room is not required to model either new BRIRs for every distance or a large BRIR database measured at different distances which would then be interpolated. This can simply be implemented by adjusting the D/R of the initial BRIR, as it seems to be the only factor from the BRIRs that affects the distance perception.

One possible source of error is the lack of training before the implementation of the test. Absence of training possibly affected more the comparison pairs of Part 1 than the ones of the other two parts of the listening test, as they were always the first tested. Furthermore, the A/B comparison and the continuous slider increased the variability across trials. Maybe, the modified MUSHRA test designed by Devalez (2008) to test relative distance would reduce the variability across trials.

As there was no headphone equalisation applied and non-individualised BRIRs used for the synthesis of the sound sources, it might be questionable if there are any issues with the externalisation, because both headphone equalisation and individualised HRTFs are important for good externalisation (section 3.2). The author's opinion is that in the current experiment externalisation is not an issue as listeners were able to discriminate sound source distances.

8. Conclusion

This project looked into the relative distance perception of virtual speech sources ranging between 1.0m and 3.0m. The virtual sources were synthesized both with measured and modelled BRIRs. The BRIRs were modelled using an image source model to design the first 80ms of the response and exponentially decaying noise for the late reverberation. The direct sound and the early reflections were rendered using MIT KEMAR HRTFs.

The aim of this project was threefold: first, to evaluate the difference between modelled and measured BRIRs in providing relative distance information; second, to investigate the effect of D/R in relative distance perception; third, to explore the effect of reflections TOA on relative distance perception.

A pairwise listening test was conducted. Subjects were asked to judge in a dimensionless scale how far or close was the second sound source of the pair in comparison to the first. Relative distance perception between virtual sources was tested under four scenarios:

- Virtual sources with normalized level
- Virtual sources whose D/R was the only factor of the BRIR that was manipulated
- Virtual sources whose reflections TOA was only manipulated
- Virtual sources without normalized level

The results from the subjective test show that:

1. Modelled BRIRs are sufficient to provide a sense of source distance and thus of externalisation, at least for the ranges studied.

2. Subjects can distinguish reliably shorter from longer distances between two sources even when level cues are absent

3. D/R is indeed a significant cue in providing relative distance perception and that its sole manipulation appears to be as effective as modelling the entire BRIR of the new source position. The implications of this finding are that D/R manipulation, which is computationally cheap, is as effective in providing distance cues as entire BRIR or room modelling, which is computationally expensive.

4. The manipulation of reflection TOA in isolation from D/R is not an effective means of providing distance cues.

9. Further work

The results from this experiment might be less applicable to other types of source signals and angular locations because it is argued that information from each distance cue is processed in different ways depending on source direction and source signal. Results from Zahorik (2002a) experiment show that the way distance cues are processed depends also on the signal type and direction of the source. For example, D/R is weighted less for speech signals than noise signals. Therefore, it would be very interesting to conduct a listening test in order to evaluate how listeners perceive relative distance using different sounds and angular locations.

Sole manipulation of BRIR's D/R is as effective as the full modelling scenario in providing distance cues for a new source location. However, the degree of realism was not examined. Therefore, it would be very interesting to evaluate how realistic is the individual manipulation of D/R compared to the full modelling scenario.

Finally, another important aspect that has to be investigated is the distance perception of amplitude panned sound sources of virtual sound systems. How can these virtual panned sources be perceived constantly in the same distance regardless of their directional location and source signal?

References

[Allen, J.B, Berkley, D.A. (1979)], "Image method for efficiently simulating small-room acoustics", *Journal of the Acoustical Society of America*, Volume 65 (4), pp. 943–950.

[Begault, D.R. (1992)], "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems", *The Journal of the Audio Engineering Society*. Volume 40 (11),pp. 895-904.

[Begault, D.R., Lee, A.S., Wenzel, E.M., Anderson, M.R. (2000)], "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualised Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source", *Audio Engineering Society 108th convention.*

[Békésy, G. V. (1949)], "The moon illusion and similar auditory phenomena", *American Journal of Psychology,* Volume *62*, pp. 540-552.

[Bejoy, J. (2002)], "Virtual Surround Sound Implementation Using Deccorelation Filters And HRTF", *Centre for computer research in music and acoustics, Stanford University.*

[Bech, S., Zacharov, N. (2006)], "Perceptual Audio Evaluation: Theory Method and Application", *WILEY*

[Blauert, J. (1976)], "Spatial Hearing" (Revised Edition), *The MIT Press*

[Bronkhorst, A. and Houtgast, T. (1999)], Auditory distance perception in rooms", *Nature*, Volume 397, pp. 517–520.

[Brookes, T. (2005)], "The effect of non-symmetrical left/right recording pinna on the perceived externalisation of binaural recordings", *Audio Engineering Society 118th Convention.*

[Brown, C.P., Duda, R.O. (1998)], "A structural model for binaural sound synthesis", *IEEE Transactions on Speech & Audio Processing,* Volume 6 (5), pp. 476-488.

[Brungart, DS. (1998)], "Control of Perceived Distance in Virtual Audio Displays", *IEEE Engineering in Medicine and Biology Society*, Volume 20 (3), pp. 1101-1104.

[Brungart, DS. (1999a)], "Auditory localisation of nearby sources I Head-related transfer functions ", *The Journal of the Acoustical Society of America*, Volume 106 (6), pp. 1465-1479.

[Brungart ,DS. (1999b)], "Auditory localisation of nearby sources. II. Localisation of a broadband source", *The Journal of the Acoustical Society of America*, Volume 106 (6), pp. 1956-1968.

[Brungart, DS. (1999c)], "Auditory localisation of nearby sources. III. Stimulus effects", *The Journal of the Acoustical Society of America'*, Volume 106 (6), pp. 3589-3602.

[Brungart, DS. (1999d)], "Auditory parallax effects in the HRTF for nearby sources", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* pp. 171-174.

[Brungart, DS, Scott, KR. (2001)], "The effects of production and presentation level on the auditory distance perception of speech", *The Journal of the Acoustical Society of America*, Volume110 (1), pp. 425-440.

[Butler, A.R., Levy, E.T., Neff, W.D. (1980)], "Apparent distance of sounds recorded in echoic and anechoic chambers", *Journal of experimental psychology Human perception and performance*, Volume 6 (4), pp. 745-50.

[Cabrera, D., Gilfillan, D. (2002)], "Auditory distance perception of speech in the presence of noise", *8th International Conference on Auditory Display*, pp. 431–439.

[Cabrera, D., Jeong, D., Kwak, H.J., Kim, J.-Y. (2005)],"Auditory Room Size Perception for Modelled and Measured Rooms", *The 2005 Internoise Congress and Exposition on Noise Control Engineering.*

[Carline, S (1996)], "Virtual Auditory Space: Generation and Applications", *R.G. Landes*

[Carty, B. (2010)], "Movements in Binaural Space : Issues in HRTF Interpolation and Reverberation , with applications to Computer Music Volume 1 of 2", *PhD Thesis NUI Maynooth Music Department.*

[Carty, B. (2010)], "Movements in Binaural Space : Issues in HRTF Interpolation and Reverberation , with applications to Computer Music Volume 2 of 2", *PhD Thesis NUI Maynooth Music Department.*

[Coleman, P.D. (1968)], "Dual Role of Frequency Spectrum in Determina- tion of Auditory Distance", *The Journal of the Acoustical Society of America*, Volume 44 (2), pp. 631-632.

[Cox, T., Antonio, P. (2009)], "Acoustic Absorbers and Diffusers: Theory, Design and Application", *CRC Press.*

[Devallez, D. (2009)], "Auditory Perspective: perception, rendering, and applications.", *Universita degli Studi di Verona Dipartimentodi Informatica.*

[Devallez, D, Fontana, F., Rocchesso, D. (2008)], "Linearizing Auditory Distance Estimates by Means of Virtual Acoustics", *Acta Acustica united with Acustica*, Volume 94 (6), pp. 813-824.

[Devore, S. (2003)], "Perceptual Concequenses of Inclusing Reverberation in Spatial Auditory Displays", *International Conference on Auditory Display.*

[Durlach, N.I., Rigopulos, A., Pang X.D., Woods, W.S., Kulkarni, A., Colburn, H.S., Wenzel, E.M. (1992)], "On the externalisation of auditory images.", *Presence Teleoperators and Virtual Enviroments*, Volume 1, pp. 251-257.

[Epstein, M, Florentine, M. (2009)], " Binaural loudness summation for speech and tones presented via earphones and loudspeakers", *Ear Hear,* Volume 30 (2) pp 234-247.

[Extra, D., Simmer, U., Fischer, S. (2006)], "Artificial Reverberation Comparing algorithms by using monaural analysis tools", *Audio Engineering Society 121st Convention.*

[Farina, A. (2000)], "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique", *Audio Engineering Society* 108^{th} Convention.

[Farina, A.], "Aurora Software", *http://www.aurora-plugins.com/.*

[Fazenda, B. (2000)], "Perception of Room Modes in Critical Listening Spaces", *PhD Thesis University of Salford.*

[Fukuda, T, Horiuchi, T, Hokari, H, Shimada, S. (2003)], "Relative distance perception by manipulating the ILD of HRTFs", *Acoustical Science and Technology*, Volume 24 (5), pp. 325-326.

[Gardner, B, Keith, M. (1994)], "HRTF Measurements of a KEMAR Dummy-Head Microphone", *MIT Media Lab - Technical Report*

[Gardner, M.B. (1968)], "Proximity Image Effect in Sound Localisation", *The Journal of neuroscience : the official journal of the Society for Neuroscience*, Volume 43(1), pp. 163

[Gardner, M.B. (1969)], "Distance estimation of 0^{o} or apparent 0^{o}-oriented speech signals in anechoic space", *The Journal of the Acoustical Society of America*, Volume 45, pp. 47–53.

[Gardner, W.G. (1992)], "Master's thesis :The virtual acoustic room", *Massachusetts Institute of Technology.*

[Gardner W.G. (1999)], "3D audio and acoustic environment modelling", The *Wave Arts 3D Technology,* Wave Arts Inc.

[Green, D.M. (1988)], "Psychoacoustics",*CHABA Symposium on sound Localisation by Humans, National Academy Sciences (unpublished)".*

[Griesinger, D. (1990)], "Binaural techniques for music reproduction", *Audio Engineering Society 8th International Conference*, pp. 197-207.

[Haas, H. (1951)],"On the influence of a single echo on the intelligibility of speech.", *Acustica,* Volume 1, pp. 48–58

[Hammershøi, D., Møller, H. (2005)], "Binaural Technique - Basic Methods for Recording, Synthesis and Reproduction", *Springer-Verlag*, pp. 379.

[Hameed, S. (2004)], "Psychoacoustic Cues in Room Size Perception", *Audio Engineering Society* 116th Convention.

[Handzel, A., Krishnaprasad, P.S. (2002)], "Biomimetic sound-source localisation", *IEEE Sensors Journal*, Volume 2 (6), pp. 607-616.

[Hartmann, WM. (1996)], "On externalisation of Auditory Images", *The Journal of the Acoustical Society of America*. Volume 99 (6), pp. 3678-3688.

[Hartmann, WM. (2011)], "How we localise sound", *Physics Today*, Volume 52 (11), pp. 24-29.

[Hirvonen, T. (2002)], "Headphone Listening Test Methods", *Master Thesis Helsinki University of Technology-Department of electrical and Communication Engineering.*

[Hofman, P.M., Van Riswick, J.G., Van Opstal, J. (1998)], "Relearning sound localisation with new ears", *Nature neuroscience*, Volume 1 (5), pp. 417-421.

[Hollander, M., Wolfe D. A. (1999)], “Nonparametric Statistical Methods”, *Hoboken, NJ: John Wiley & Sons, Inc.*

[Howard, D.M., Angus, J. (2009)],"Acoustics and Psychoacoustics 4th edition", *Focal Press.*

[Huopaniemi, J., Karjalainen, M. (1996)], "Comparison of Digital Filter Design Methods for 3-D Sound", *IEEE Nordic Signal Processing Symposium.*

[Ihlefeld, A., Shinn‐Cunningham, B.G. (2004)], "Effect of source location and listener location on ILD cues in a reverberant room", *The Journal of the Acoustical Society of America*, Volume 115 (5), pp. 2598-2598.

[Jot, J-M, Larcher, V., Warusfel, O. (1995)], "Digital Signal Processing Issues in the Contex Of Binaural And Transaural Stereophony", *Audio Engineering Society 98th Convention.*

[Jot, J-M. (1997)], "Efficient models for reverberation and distance rendering in computer music and virtual audio reality", *International Computer Music Conference.*

[Kapralos, B. (2006)], "The Sonel Mapping Acoustical Modelling Method" *19th Iinternational Congress on Aacoustics.*

[Kendall, G.S. (1995)], "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery", *Computer Music Journal*, Volume 19(4), pp. 71-87.

[Kim, H-Y, Suzuki, Y, Takane, S, Sone, T. (2001)] "Control of auditory distance perception based on the auditory parallax model", *Applied Acoustics*. Volume 62 (3), pp. 245-270.

[Kim, S-M, Choi, W. (2005)], "On the externalisation of virtual sound images in headphone reproduction: A Wiener filter approach.", *The Journal of the Acoustical Society of America*, Volume 117 (6), pp. 3657:265

[Kim, Y.G., Chun, C.J., Kim, H.K., Lee, Y.J. (2010)], "An Integrated Approach of 3D Sound Rendering Techniques for Sound Externalisation.", *Advances in Multimedia Information Processing*, Volume 1, pp. 682-693.

[Kopčo, N. (2003)], "Spatial hearing, auditory sensitivity and pattern recognition in noisy environments,", *Ph.D. Thesis, Boston University.*

[Kopčo, N, Čeljuska, D, Puszta, M. (2004)], "Effect of spectral content and learning on auditory distance perception." *2nd Slovak-Hungarian Joint Symposium on Applied Machine Intelligence* .

[Kopčo, N, Shinn-Cunningham, B.G. (2011)], "Effect of stimulus spectrum on distance perception for nearby sources.", *The Journal of the Acoustical Society of America*. Volume 130 (3), pp.1530-1541.

[Larcher, V., Jot J.M., Vandernoot, G. (1998)], "Equalisation Methods In Binaural Technology", *Audio Engineering Society 105th Convention*

[Larsen, E, Iyer, N, Lansing, C.R., Feng, A.S. (2008)], "On the minimum audible difference in direct-to-reverberant energy ratio." *The Journal of the Acoustical Society of America*, Volume 124 (1), pp. 450–461.

[Langendijk, E.H., Bronkhorst, A.W. (2000)], "Fidelity of three-dimensional-sound reproduction using a virtual auditory display", *The Journal of the Acoustical Society of America*, Volume 107 (1), pp. 528-537.

[Lienard J. S., Benedetto, M.G.D. (1999)], "Effect of vocal effort on spectral properties of vowels", *The Journal of the Acoustical Society of America,* Volume 106, pp. 411–422.

[Loomis, J.M., Herbert, C., Cincinelli, J.G. (1990)], "Active localisation of virtual sounds", *The Journal of the Acoustical Society of America*, Volume 88, pp.1757-1764.

[Lu, Yun-Chen 2010], "Active Hearing Strategies for Binaural Sound Localisation in Azimuth and Distance by Mobile Listeners.", *PhD* Thesis *Department of Computer Science University of Sheffield*.

[Mathworks Matlab (2011)],"'DSPToolbox*", http://www.mathworks.co.uk.*

[Martens, WL. (2003)], "Perceptual evaluation of filters controlling source direction : Customized and generalized HRTFs for binaural synthesis.", *Acousic. Science & Technology, Volume 24,* pp. 220-232.

[Matigue J.J. (2001)], "Impact of Artificial Reverberation on Perceived Sound Localisation during a Headphone Listening Task", *Audio Engineering Society 11th Convention.*

[McKeag, A., McGrath, D. (1997)], "Using auralisation techniques to render 5.1 surround to binaural and Transaural playback", *Audio engineering Society 102nd Convention.*

[Menzer, F. (2010)], "Binaural reverberation using two parallel feedback delay networks.", *Audio Engineering Society.*

[Mershon, D.H., King, L.E. (1975)], "Intensity and reverberation as factors in the auditory perception of egocentric distance.", *Perception*, Volume 18 (6), pp. 409-415.

[Mershon, D.H., Bowers, J.N. (1979)], "Absolute and relative cues for the auditory perception of egocentric distance.", *Perception*, Volume 8 (3), pp. 311-322.

[Mershon, D.H., Desaulniers, D.H., Amerson, T.L., Kiefer S.A. (1980)], "Visual capture in auditory distance perception: proximity image effect reconsidered," *The Journal of Auditory Research*, Volume 20, pp. 129–136

[Mershon, D.H., Philbeck, J.W. (1998)], "Auditory perceived Distance of familiar speech sounds", *Proceedings of the Psychonomic Society 32nd Annual Meeting.*

[Michelsen, J. (1997)], "Parameters of distance perception in stereo loudspeaker scenario.", *Audio Engineering Society 102nd Convention*

[Middlebrooks, J.C., Green, DM. (1991)], "Sound localisation by human listeners. *Annual Review of Psychology*", Volume 42 (1), pp. 135-159.

[Møller, H., Hammershøi, D., Jensen, C.B., Sørensen, M. F. (1995)], "Head Related Transfer Functions of Human Subjects", *Journal of the Audio Engineering Society* Volume 43, pp. 300-321.

[Møller, H., Sorensen, M. F., Jensen, C. B., and Hammershari, D. (1996)], "Binaural technique: do we need individual recordings?", *Journal of the Audio Engineering Society,* Volume 44, pp. 451-469.

[Morein-Zamir, S., Soto-Faraco, S., Kingstone, A. (2003)], "Auditory capture of vision: examining temporal ventriloquism", *Cognitive Brain Research*, Volume 17, pp.154–163.

[Moore, H.A., Tew, A., Nociol, R. (2007)], "Headphone transparification : A novel method for investigating the externalisation of binaural sounds." *Audio Engineering Society 123rd Convention.*

[Morimoto, Y., Nishino, T., Takeda, K. (2010)], "Visualization and dereverberation of head-related transfer function based on spatio-temporal frequency analysis." *International Congress on Acoustics.*

[Nielsen S.H. (1993)], "Auditory Distance Perception in Different Rooms *", *Journal of Audio Engineering Society,* Volume 41 (10), pp. 755-770.

[Ohnishi, N, Sugie, N. (1998)], "Spatial localisation of sound sources: azimuth and elevation estimation", *IEEE Instrumentation and Measurement Technology Conference. Where Instrumentation is Going*, pp. 330-333.

[Oppenheim, A.V, Verghese, G.C. (2010)], "Signals, Systems, and Interference", MIT: *Class Notes for 6.011: Introduction to Comunication, Control and Signal*

[Paquier, M., Koehl, V. (2010)], "Audibility of headphone positioning variability" *Audio Engineering Society 128th Convention.*

[Park, Y, Hwang, S, Park, Y.S. (2009)], "A probabilistic method for analysis of sound localisation performance", *Applied Acoustics*, Volume 70 (5), pp. 771-776.

[Pellegrini, R.S. (2001)], "A virtual reference listening room as an application of auditory virtual environments*", Dissertation IKA, Ruhr-University of Bohum, Germany.*

[Pellegrini, R.S. (2002)], "Perception-Based Design of Virtual Rooms For Sound Reproduction", *Audio Engineering Society 22nd International Conference on VSEA.*

[Philbeck, J.W., Mershon, D.H. (2002)], "Knowledge about typical source output influences perceived auditory distance.", *The Journal of the Acoustical Society of America*, Volume 111 (5), 1980-1983.

[Picinali, L. (2010)], "The Creation of a Binaural Spatialization Tool", *PhD Thesis De Montfort University.*

[Plenge, G. (1974)], "On the differences between localisation and lateralization then stressed." *The Journal of Audio Engineering Society of America*, Volume 56 (3), pp. 944-951.

[Qu, T., Xiao, Z., Gong, M. (2009)], "Distance-Dependent Head-Related Transfer Functions Measured With High Spatial Resolution Using a Spark Gap", *IEEE Transactions on Audio, Speech, and Language Processing*, Volume 17 (6), pp. 1124-1132.

[Rychtáriková, M., Bogaert, T., Vermeir, G., Wouters, J. (2009)], "Binaural sound source localisation in real and virtual rooms.", *The Journal of the Acoustical Society of America*, Volume 57 (4), pp. 205-220.

[Strybel, T.Z, Perrott, D.R. (1984)], "Discrimination of relative distance in the auditory modality: the success and failure of the loudness discrimination hypothesis." *The Journal of the Acoustical Society of America*. Volume 76 (1), pp. 318-320.

[Sakamoto, N, Gotoh, T, Kimura, Y. (1975)], "On Out-of-Head Headphone Localisation Listening", *Journal of the Audio Engineering Society*, Volume 24 (9), pp. 710-716.

[Santarelli, S.G. (2001)], "Auditory Localisation of Nearby Sources in Anechoic and Reverberant Environments", *PhD Thesis Boston University.*

[Schoolmaster M.; Kopčo, N., Shinn-Cunningham, B.G. (2003a)], "Effects of reverberation and experience on distance perception in simulated environments", *Journal of Acoustical Society of America*, Volume 113 (4) , pp. 2285-2285.

[Schoolmaster, M, Kopčo, N, Shinn-cunningham, B.G. (2003b)], "Auditory Distance Perception in Fixed and Varying Simulated Acoustic Environments" *Journal of Acoustical Society of America,* Volume 11 , pp. 2459.

[Schroeder, M.R. (1979)], "Binaural dissimilarity and optimum ceilings for concert halls : More lateral sound diffusion", *The Journal of the Acoustical Society of America*, Volume 65 (4), page 958-963.

[Sheeline, C.W. (1982)], "An investigation of the effects of direct and reverberant signal interactions on auditory distance perception", *PhD Thesis Stanford University Department of Hearing and Speech.*

[Shinn-Cunningham, B.G. (2000a)], "Learning Reverberation : Considerations for Spatial Auditory Displays", *International Conference on Auditory Display*, pp. 126-134.

[Shinn-Cunningham, B.G. (2000b)], "Distance Cues for Virtual auditory Space", *IEEE-PCM Special session on Vrtual Auditory Space.*

[Shinn-Cunningham, B.G. (2000c)], "Distance perception of nearby sources in reverberant and anechoic listening conditions", *23rd mid-Winter meeting of the Association for Research in Otolaryngology.*

[Shinn-Cunningham, B.G., Santarelli, S., and Kopčo, N. (2000d)], "Tori of confusion: Binaural localisation cues for sources within reach of a listener", *Journal of the Acoustical Society of America*, Volume 107, pp. 1627-1636.

[Shinn-Cunningham, B. (2001)], "Creating three dimensions in virtual auditory displays", *HCI International.*

[Shinn-Cunningham, B. (2003)], "Acoustics and Perception of Sound in Everyday environments", *3rd International Workshop on spatial Media.*

[Shinn-Cunningham, B.G., Kopčo, N, Martin, T.J. (2005)], " Localising nearby sound sources in a classroom: binaural room impulse responses", *The Journal of the Acoustical Society of America,* Volume 117 (5), pp. 3100-3115.

[Simpson, W.E., Stanton, L.D. (1973)], "Head movement does not facilitate perception of the distance of a source of sound", American Journal of Psychology, Volume 86, pp. 151–159.

[Smyth, S., Smyth, M., Cheung, S. (2008)], "Headphone Surround Monitoring For Studios.", *Audio Engineering Society 23rd UK Conference*

[Smyth, M. (2010)], "Bringing theatre sound to the desktop" *Proceedings of the Institute of Acoustics*, Volume 32 (5).

[Spors, S., Wierstorf, H., Ahrens, J. (2011)], "Interpolation and Range Extrapolation of Head-Related Transfer Functions Using Virtual Local Wave Field Synthesis", *Audio Engineering Society 130th Convention*

[Strybel, T.Z., Perrott, D.R. (1984)], "Discrimination of relative distance in auditory modality:the succuess and failure of the loudness discrimination hypothesis", *The Journal of Acoustical Society of America,* Volume 76 (1), pp. 318-20.

[Supper, B. (2010)], "Processing and improving a head-related impulse response database for auralization", *Audio Engineering Society 129th Convention.*

[Valente, D.L., Braasch, J. (2010)], "Subjective scaling of spatial room acoustic parameters influenced by visual environmental cues" *The Journal of the Acoustical Society of America*, Volume 128 (4), pp. 1952-1964.

[Van Wanrooij, M.N., Van Opstal, J. (2005)], "Relearning sound localisation with a new ear", *The Journal of neuroscience : the official journal of the Society for Neuroscience*, Volume 25 (22), pp. 5413-5424.

[Vanderkooy, J. (1994)], "Aspects of MLS Measuring Systems", *Journal of Audio Engineering Society*, Volume 42, pp. 219-231.

[Völk, F., Heinemann, F., Fastl, H. (2008)], "Externalisation in binaural synthesis: effects of recording environment and measurement procedure.," *Euronoise*, pp. 6421-6426.

[Völk, F. (2009)], "Externalisation in data-based Binaural Synthesis : Effects of Impulse Response Length", *NAG/DAGA*, pp. 1075-1078.

[Vorländer, M. (2007)], "Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality", *Springer.*

[Wallach, H. (1939)], "The two factors governing sound localisation", *The Journal of the Acoustical Society of America*, Volume 10, pp. 270-274.

[Warren, R.M. (1973)], "Anomalous loudness function for speech", *The Journal of the Acoustical Society of America* Volume 54 (2), pp. 390-396.

[Weinrich, S. (1992)], "Improved Externalisation and Frontal Perception of Headphone Signals", *Audio Engineering Society 92nd Convention.*

[Wersenyi, G. (2008)], "Effect of Emulated Head-Tracking for Reducing Localisation Errors in Virtual Audio Simulation", *IEEE Transactions on Audio, Speech, and Language Processing.*

[Wightman, F.L., Kistler D.J. (1999)], "Resolution of front-back ambiguity in spatial hearing by listener and source movement", *The Journal of the Acoustical Society of America*, Volume 105 (5), pp. 2841-2853.

[Wilslon, K., (2005)], "Learning the Precedence effect", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.*

[Zahorik, P. (1997)], " Scaling perceived distance of virtual sound sources", *Acoustical Society*

[Zahorik, P. (2000), "Distance localisation using nonindividualized head-related transfer functions." *The Journal of the Acoustical Society of America*, Volume 108 (5), pp. 2580-2631.

[Zahorik, P. (2001)], "Estimating sound source distance with and without vision", *Optometry and vision science*", Volume 78 (5), pp. 270-2755.

[Zahorik, P. (2002a)], "Assessing auditory distance perception using virtual acoustics", *The Journal of the Acoustical Society of America*. Volume 111 (4), pp. 1832-1846.

[Zahorik, P. (2002b)], "Auditory Display of Sound Source Distance", *International Conference on Auditory Display.*

[Zahorik, P. (2002c)], "Direct-to-reverberant energy ratio sensitivity", *The Journal of the Acoustical Society of America*, Volume 112 (5), pp. 2110-2117.

[Zahorik, P, Brungart, D.S., Bronkhorst, A.W. (2005)],"Auditory Distance Perception in Humans : A Summary of Past and Present Research" *Acta Acustica united with Acustica*, Volume 91 pp. 409 - 420.

APPENDIX

APPENDIX A

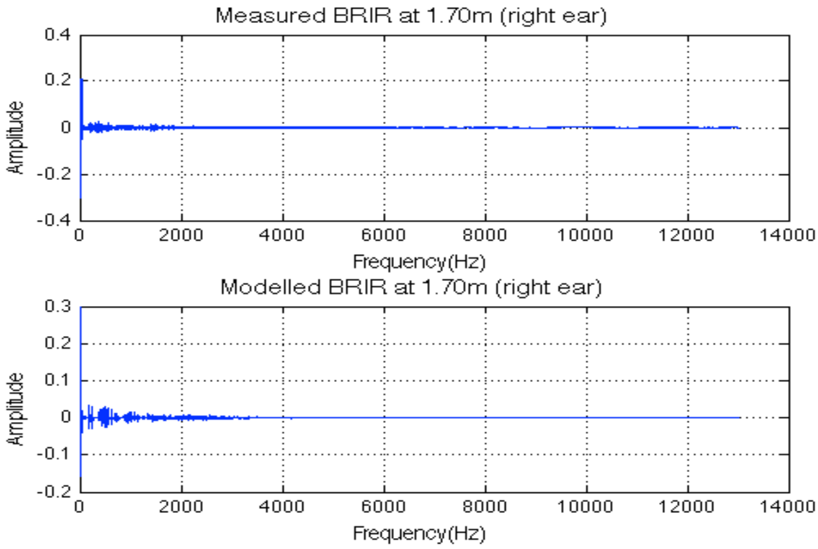


Figure A.1: Listening room measured and modelled right ear BRIR at 1.70m and 45° azimuth

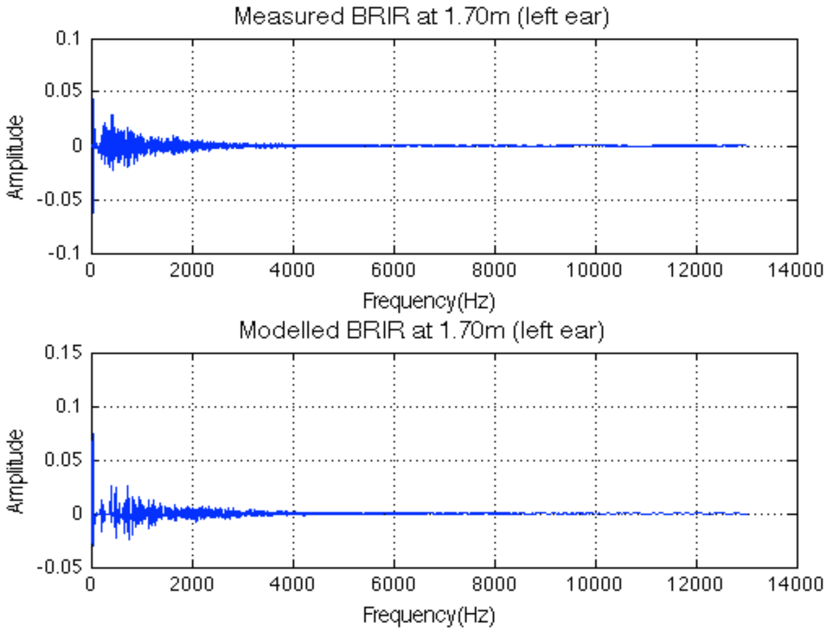


Figure A.2: Listening room measured and modelled left ear BRIR at 1.70m and 45° azimuth

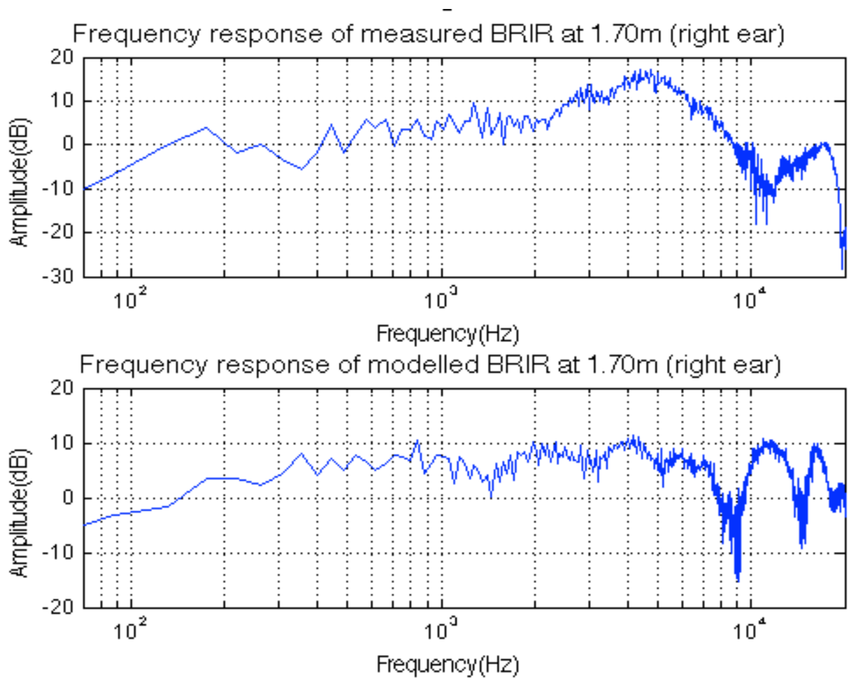


Figure A.3: Frequency response of measured and modelled right ear BRIR of the listening room at 170m and 45° azimuth

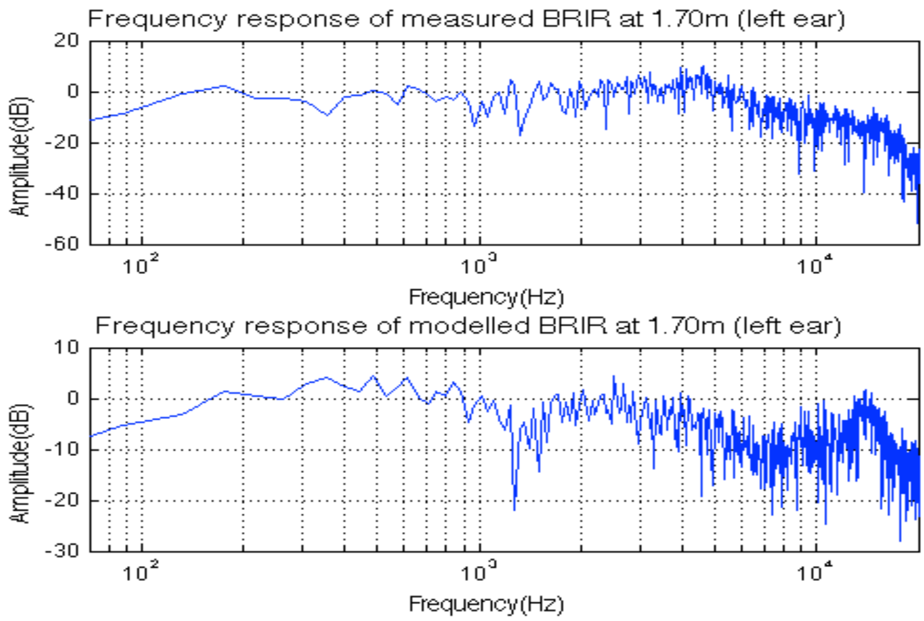


Figure A.4: Frequency response of measured and modelled left ear BRIR of the listening room at 170m and 45° azimuth

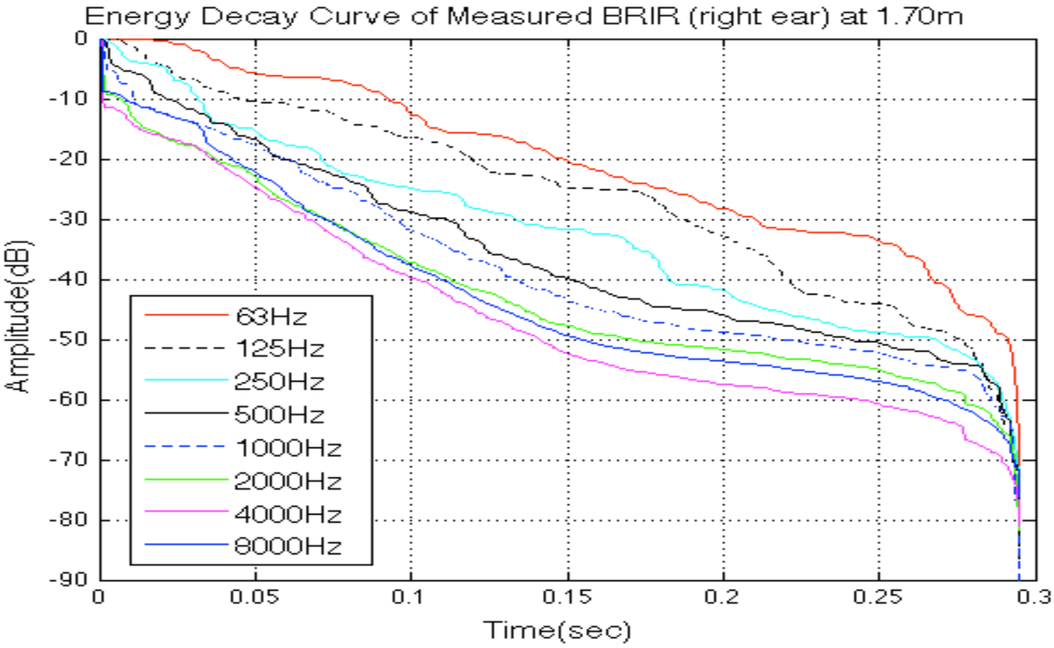


Figure A.5: Energy decay curves of right ear BRIR measured at the listening room at 1.70m and at 45°

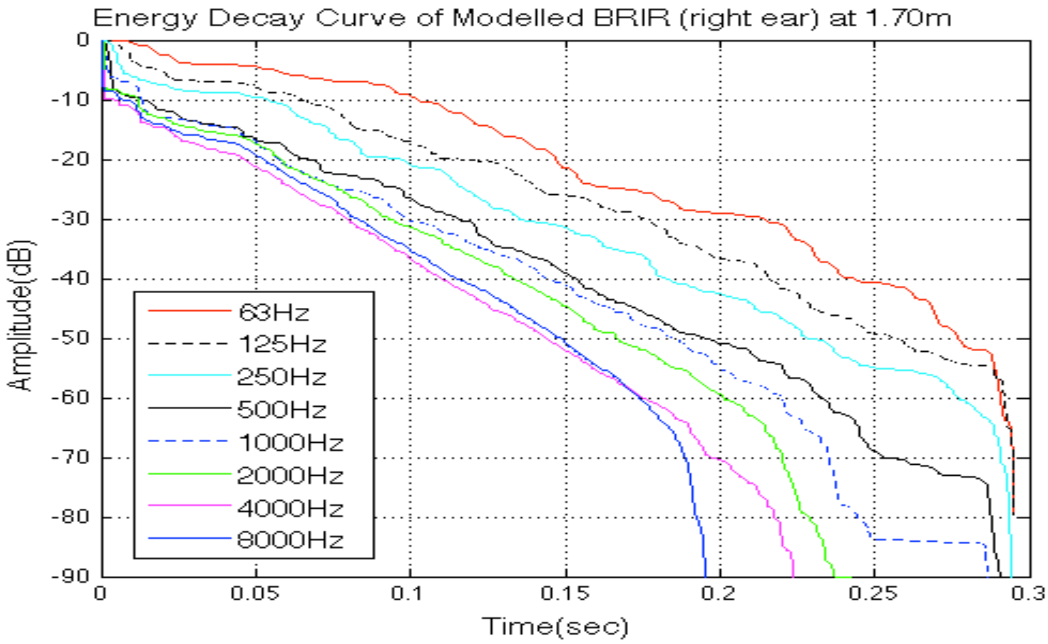


Figure A.6: Energy decay curves of right ear BRIR modelled at the listening room at 1.70m and at 45°

APPENDIX B

Listening tests for distance perception

General information

There will be three tests and the overall time will be approximately 25 minutes.

For all tests you are required to listen to different pairs of sounds. You will have to note whether the second sound is at a distance farther **or** closer compared to the first sound. **Please note that some of the cues relating to distance might be very subtle so I ask you to listen critically to anything that might indicate a distance difference.** Before starting test 1 you will listen to all the samples a few times in order to appreciate the context of the test. The same will happen before test 3.

Important notes:

- The evaluation will be done through a continuous slider
- **The first sound of the pairs is always the reference sound. So, with the slider you are judging the distance perception of "sound 2" in comparison to the reference sound and not the other way round**.
- Between reference and "sound 2" there will be a gap of 0.2 seconds
- Before moving to the next pair, always confirm your answer on the tick box and then press next.
- You can play the same pair of sounds as many times as you want, but a couple of times should be enough to come to a decision
- The sources are always presented at the same angular location (right-hand side). You don't have to worry about that. Concentrate only on distance.
- Try to keep your head stable during the playback.
- It is suggested to keep you eyes closed during the playback because it can help you concentrate better
- When you finish one test, I will load the next test window.
- If you feel tired, feel free to have a break